Universidade de Aveiro
2011

Departamento de Química

Leonildo Pedro Cabral Delgado

Estudos estruturais da versão humana da proteína p22HBP

Structural studies of human p22HBP protein

Universidade de Aveiro
2011

Departamento de Química

Leonildo Pedro Cabral Delgado

Estudos estruturais da versão humana da proteína p22HBP

Structural studies of human p22HBP protein

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biotecnologia, realizada sob a orientação científica do Doutor Brian James Goodfellow, Professor Auxiliar do Departamento de Química da Universidade de Aveiro

Dedico este trabalho à memória dos meus Pais.

o júri

presidente

Prof. Doutor Jorge Manuel Alexandre Saraiva
Investigador auxiliar do Departamento de Química da Universidade de Aveio

Prof. Doutor Brian James Goodfellow
professor auxiliar do Departamento de Química da Universidade de Aveiro

Prof. Doutora Maria dos Anjos López Macedo
professora auxiliar do Departamento de Química da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

agradecimentos

Agradeço a todas as pessoas que contribuiram para a realização deste trabalho.

Ao meu orientador, Doutor Brian, por me ter acolhido e mostrado este mundo fascinante das proteínas e RMN.

À Susana, pela paciência demonstrada e todos os conhecimentos transmitidos.

Ao Cláudio, à Joana e ao Tiago, por todo o companheirismo e disponibilidade ao longo deste ano lectivo.

Ao grupo de espectrometria de massa, pela disponibilização de todos os equipamentos que me foram de grande utilidade.

À minha família, por todo o apoio.

A ti Ana, por fazeres parte da minha vida e teres estado sempre presente, nos melhores e piores momentos, ao longo deste ano lectivo, contribuindo de forma positiva para a realização desta tese.

A todos, um "muito obrigado" por me terem acompanhado ao longo deste percurso.

palavras-chave

Proteínas recombinantes, sobrexpressão, purificação, RMN, porfirinas, fluorescência

resumo

O presente trabalho teve por objectivos realizar os primeiros estudos estruturais da versão humana da proteína p22HBP por RMN e ainda estudar a interacção de várias porfirinas com esta proteína usando a extinção da fluorescência intrínseca dos resíduos de triptofano. A realização destes objectivos pressupõe a optimização das condiçoes de produção e purificação da p22HBP humana, tendo-se testado para isso parâmetros com influência no total de proteína produzida e purificada. Esses parâmetros são o momemto da indução, a concentração do agente indutor e o tempo de lise celular por sonicação. Após a optimização das condições referidas, a produção da p22HBP tem lugar num meio de cultura contendo isótopos como $^{15}NH_4Cl$ e ^{13}C-glucose, de modo a incorporar nas proteínas esses núcleos activos em RMN. O passo seguinte consistiu na identificação das ressonâncias da cadeia principal da p22HBP humana, primeiro passo e determinante na elucidação da estrutura tridimensional de qualquer proteína, por RMN. Por fim, foram realizados estudos de interacção usando o mapeamento das diferenças de desvio químico devidas à interacção com as porfirinas e, ainda a fluorescência intrínseca dos resíduos de triptofano para determinar a afinidade da proteína em relação às porfirinas. Estes estudos indicaram que o sítio de interacção presente na proteína é composta pelos mesmos resíduos previamente identificados no estudo com a murine HBP; para além disso, a constante de dissociação para as porfirinas situa-se na gama dos nanomolares, havendo portanto uma elevada afinidade.

keywords

Recombinant protein, overexpression, purification, NMR, porphyrins, fluorescence

abstract

The present work aimed to perform the first structural studies on human p22HBP protein by NMR and at same time to study the interaction of some tetrapyrrole with this protein by fluorescence quenching. To accomplish these objectives we had to start by optimizing the conditions for producing and purifying human p22HBP. Parameters such as the moment of induction, the inducer concentration and time for cell lysis were evaluated. After the optimization, cultures of p22HBP in M9 minimal media enriched with $^{15}NH_4Cl$ and ^{13}C-glucose were grown. The first spectra of human p22HBP and the backbone resonance assignment, first and most important step in any protein structural elucidation by NMR, were performed. Finally, interaction studies were performed using chemical shift mapping and fluorescence quenching methods, and they indicated that the tetrapyrrole binding site is formed by the same residues described in previous studies; and the dissociation constant by which human p22 HBP binds PPIX and hemin was determined as being in the nanomolar range, being this interaction of high affinity.

List of figures

Abbreviations

Ala	Alanine
ALAS	δ-aminolevulinate synthase
APS	Ammonium persulfate
Asp	Asparagine
BAC	Bacterial artificial chromosome
BMRB	Biological magnetic resonance bank
COSY	Correlation spectroscopy
CSA	Chemical shift anisotropy
ckSOUL	Chicken SOUL
CYP450	Cytochrome P450
DD	Dipole-dipole
DNA	Deoxyribonucleic acid
EDTA	Ethylenediamine tetracetic acid
E. coli	Escherichia coli
FABP	Fatty-acid-binding protein
Gly	Glycine
gor	Glutathione reductase
GST	Glutathione S-transferase
HAC	Human artificial chromosome
HBP	Heme binding protein
His-tag	Histidine-tagg
HO	Heme oxygenase
HSQC	Heteronuclear single quantum coherence
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
Kd	Dissociation constant
lacY1	*lac* permease
LB	Luria-Bertani
LC-MS	Liquid chromatography–mass spectrometry

mHBP	Murine HBP
mSOUL	Murine SOUL
MALDI-TOF	Matrix-assisted laser desorption/ionization – Time of flight
Met	Methionine
NCBI	National center for biotechnology information
Ni-NTA	Niquel-nitrilotriacetic acid
NMR	Nuclear magnetic resonance
NOESY	nuclear Overhauser effect spectroscopy
Phe	Phenilalanine
PPIX	Protoporphyrin IX
rDNA	Recombinant DNA
RNA	Ribonucleic acid
ROS	Reactive oxygen species
SDS-PAGE	Sodium dodecyl sulphate polyacrilamide gel electrophoresis
Ser	Serine
TEMED	N,N,N′,N′-Tetramethyl ethylenediamine
TOCSY	total correlation spectroscopy
TROSY	Trnasverse optimized relaxation spectroscopy
Trp	Tryptophan
trxB	Thioredoxin reductase
Thr	Threonine
UV-Vis	Ultraviolet-visible
Val	Valine
YAC	Yeast artificial chromosome

Content

Chapter 1 – Literature Review

1.1. Heme

Heme is a tetrapyrrole with important biological roles and serves as a prosthetic group when linked covalently or non-covalently to specific hemoproteins in which it is the biological active center. It is essential for the transport of oxygen, for the production of cellular energy, in steroid biosynthesis or in signal transduction. In mammals heme is mainly utilized for hemoglobin synthesis in erythrocytes, myoglobin in muscle cells and for CYP450 in the liver. Despite being ubiquitous heme can be toxic when found in a free form due to the generation of ROS, which produces DNA damage, lipid peroxidation and protein denaturation [1].

Heme (iron protoporphyrin IX) is a tetrapyrrole with a central iron ion that is synthesized by enzymes localized in two cellular compartments. The first and the last three enzymes from this biosynthetic pathway are localized in the mitochondrion, whereas the others are localized in the cytosol (**Figure 1**). The rate limiting step of heme biosynthesis is the formation of 5-aminolevulinic acid from glycine and succinyl CoA by ALAS. The last step is catalyzed by ferrochelatase, where ferrous is inserted into protoporphyrin IX. The quantitative expression and regulation of heme synthesis is different in various tissues and is mainly regulated by ALAS activity. ALAS have two distinct isoforms, an erythroid specific isozyme (ALAS2), and a non-specific housekeeping isozyme, which is expressed in the liver and other non-erythroid tissues (ALAS1).

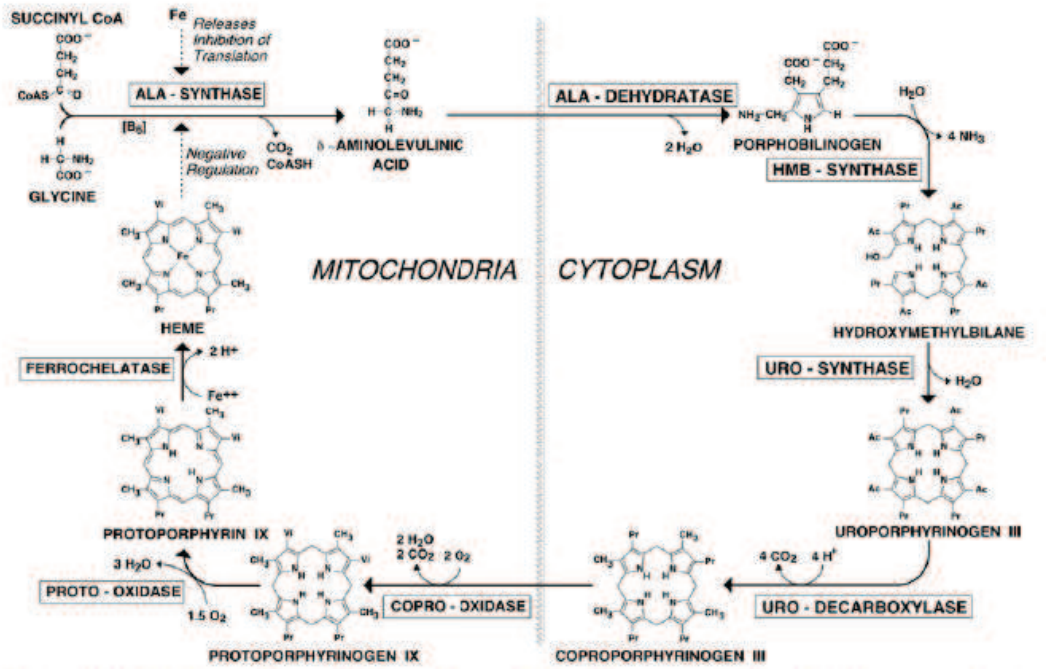

Figure 1: The heme biosynthetic pathway. The pathway uses eight enzymes, four localized in the cytoplasm and four in the mitochondria [2].

Heme synthesis takes place mostly in developing red blood cells in the marrow but about 15% of the daily production occurs in the liver for the formation of heme-containing enzymes. However the regulatory mechanisms by which these two organs control heme synthesis are different. In the liver heme biosynthetic enzymes are turned over rapidly, enabling the liver to respond to changing metabolic requirements, in erythroid progenitors the pathway is regulated in order to allow a high steady-state of heme synthesis with regulation tied to the availability of iron [3]. Adequate levels of heme are maintained through a combination of synthetic (uptake of extracellular heme into hepatocytes), and degradative mechanisms (via HO). Cells must possess a sensing system that allows them to monitor changes in the size of the uncommitted heme pool. Such system has not been identified but it seems that it may be linked to cytosolic heme-binding proteins because heme is toxic when in a non-bound form, poorly soluble in aqueous solutions at body pH and tends to form large aggregates at concentrations as low as 10^{-7} M. Several possible heme-binding and heme-transport proteins from the cytosol (FABP, GSTs and HBP23) have been documented [4].

1.2. The SOUL/HBP family

In 1998 Taketani *et al.* reported a new heme-binding protein, expressed in high levels in various tissues with the highest levels having been verified in the liver, spleen and kidney. In addition it possessed a molecular mass of 22 kDa, the reason why it was called p22HBP. Studies of this protein after over-expression by molecular biology techniques showed that it binds heme, protoporphyrin IX and coproporphyrin with relatively high affinities [4]. Later, in an attempt to identify genes expressed in retina and pineal gland in domestic chicken (Gallus gallus), a ckSoul gene was found, highly expressed in these organs [5]. The ckSOUL sequence was used to find homologs in mammals and this search returned HBP. Sequence comparisons indicate a high (87%) homology between mouse and human HBP proteins and much lower (24-28%) homology between ckSoul and mHBP and mSOUL (**Figure 2**). This indicates that mHBP, mSOUL and ckSOUL represent members of a new protein family (the SOUL/HBP family).

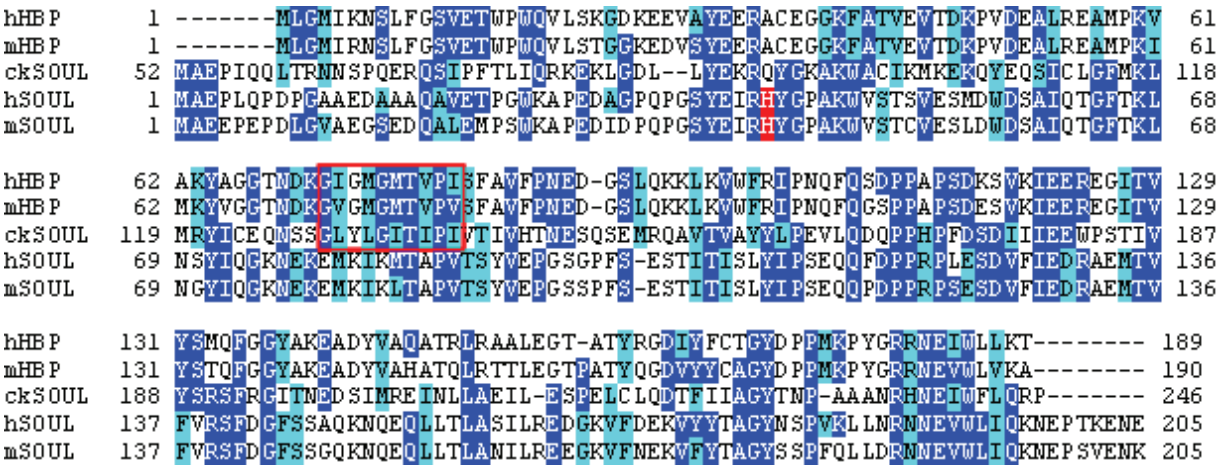

Figure 2: Protein sequence alignment between p22 HBP, ckSOUL and mammalian SOUL using CLUSTALW. Identical amino acids are shaded in dark blue and similar residues are shaded in bright blue. The p22 HBP putative heme binding hydrophobic region (common to ckSOUL but not for mammalian SOUL) is delimited in red [4-5]. Proposed His residue that axially binds the heme iron in mammalian SOUL is highlighted in red [6].

In 2002 Blackmon *et al.* reported Kd values ranging the μM, for p22HBP upon binding heme and other tetrapyrroles (hemin and protoporphyrins), but it´s role in the cell have remained unknown until 2005 [7]. Babusiak *et al.* carried out a study in which isotopically labeled [^{59}Fe]hemin was introduced into murine erythroleukemia cells and then the cellular lysates were separated by ion-exchange chromatography and the fractions containing radioactivity were collected and separated by native electrophoresis. The composition of the labeled protein complexes was analyzed by LC-MS/MS and in this way p22HBP was identified as a component in one of four multi-proteic complex associated with hemoglobin biosynthesis. They suggested that this protein can act as heme transporter or chaperone in insertion of heme into hemoglobin [8].

In 2006, the three-dimensional structure (**Figure 3**) of murine p22HB, the first for a protein from SOUL/HBP family was determined by NMR to consist of a 9-stranded distorted β-barrel flanked by two long α-helices [9]. Furthermore, the tetrapyrrole binding site was also determined by chemical shift mapping to a hydrophobic cleft formed by residues from helix α_A and an extended loop (**Figure 4**).

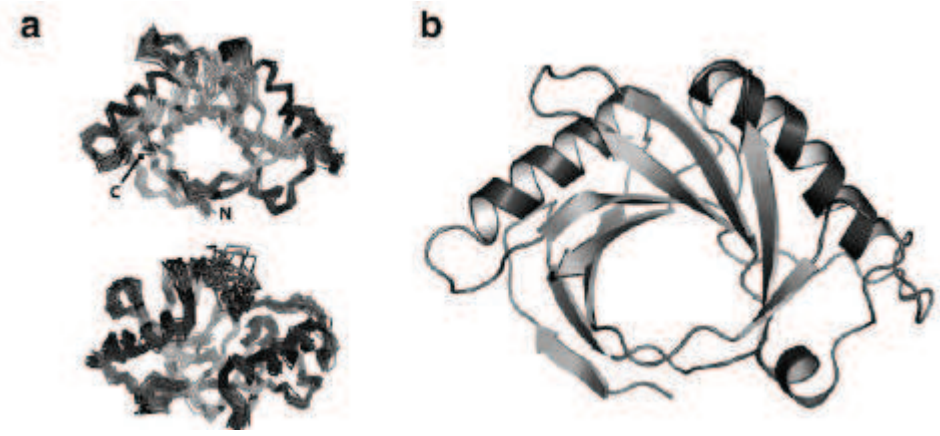

Figure 3: Resolved structure by NMR of murine p22HBP. a: ensemble of 20 conformers; b: global view of p22HBP structure [9].

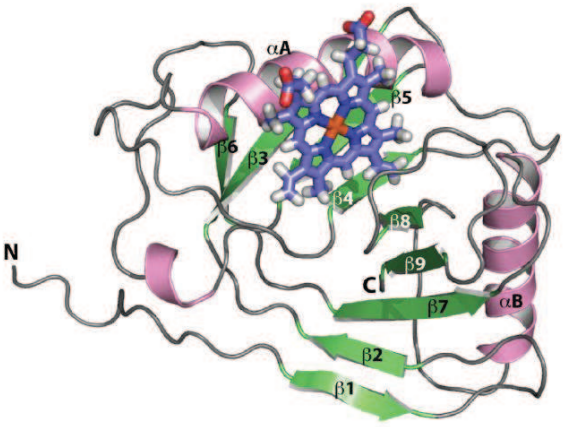

Figure 4: Model of the p22HBP-hemin complex [9].

Recently, a molecular modeling study has been carried out in order to characterize the molecular recognition process that takes place upon the binding of HBP with intermediates of heme synthesis (i.e. hemin, PPIX). This study concluded that hemin and PPIX have identical binding orientations, in which the stabilization of the propionate side chains is mainly achieved by electrostatic interactions with lysine 177 (position 176 in human) and arginine 56 located at the edges of the protein-binding pocket (**Figure 5**). Furthermore, the binding pocket of the human homologue is almost identical to the murine protein. The sequentially (and structurally) conserved lysine and arginine residues seem to play an identical role in stabilizing the tetrapyrrole propionate side chains in both proteins[10].

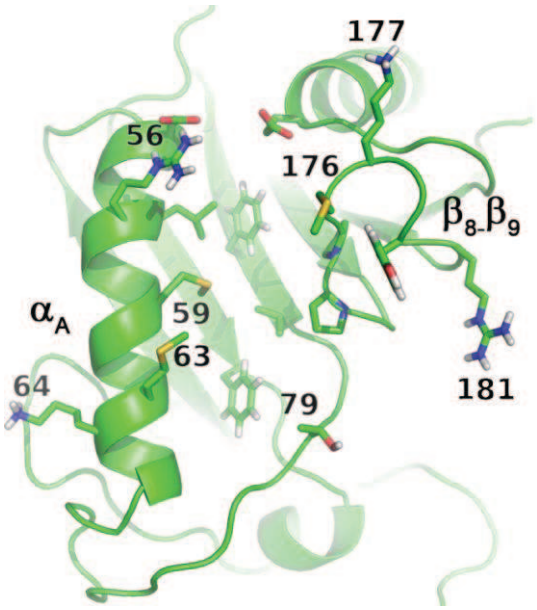

Figure 5: Cross-eye stereo view of the apo mHBP protein-binding site with selected side chain residues of the binding pocket rendered in sticks[10].

1.3. Aims

Everyday new proteins are found and the necessity for understanding their function grows. Human proteins have a special interest because knowledge of their roles in the cell may lead to the comprehension of metabolic processes and diseases. To obtain sufficient quantities of these proteins molecular biology techniques are required which results not only in sufficient amount but also adequate levels of purity for subsequent studies of structure and function.

The comprehension of the structure and function of the proteins that interact with heme is essential to fully understand the transport and regulation of heme in biological systems. So far, the proteins of the HBP/SOUL family studied have been from a murine source (mHBP), so now, as a more relevant model for future mutagenesis studies, the first structural and interaction studies in the human version of the HBP is being attempted and the comparation with murine as well. As specific ally we want to:

1. Optimize the conditions for over-expression and purification of human version of p22HBP protein.

2. Assign the backbone resonances of human p22HBP.

3. Study the interaction of some porphyrins with this protein by chemical shift mapping and fluorescence quenching.

Chapter 2 – Introduction

2.1. Recombinant protein production

One of the major challenges in large protein structural studies is the production of correctly folded protein and in enough quantities. Protein over-expression refers to the directed synthesis of large amounts of a desired protein and production of heterologous proteins involving the introduction of foreign DNA into the host cells [11]. In the 1970´s new molecular biology techniques were developed allowing isolation and purification of target genes in a process called genetic cloning. Recombinant DNA technology, as these techniques are called, can be used to study mechanisms of replication and genetic expression, to determine gene sequences and ultimately the sequence of protein it encodes, and for development of cultures capable of producing pharmacologically active products [12]. The first reported production of human recombinant protein took place during the 70's by the biotech company Genentech that expressed the gene encoding somatostatin in *E*. coli. Later, in 1982, this company followed this success by producing humulin, recombinant insulin produced in *E*. coli and the first biotech drug to be accepted for market by the Food and Drug Administration [13-14]. For production of any heterologous proteins one must be aware of some considerations: identification and isolation of the DNA to be introduced; identification of the vector and construction of recombinant vector; and identification of the suitable expression system to receive rDNA [15].

In the beginning of the recombinant protein expression era the bacteria *E.* coli and Bacillus spp. were almost the only hosts that were used for recombinant expression. But the knowledge that a protein may require a specific host physiology and biochemistry to achieve optimal production has fired up the search for new hosts. There are two major categories of expression systems available, prokaryotic and eukaryotic. Prokaryotic systems are generally the most used but they have many limitations when using the cells for the production of eukaryotic proteins, e.g. many eukaryotic proteins undergo post-translational modifications. However there is no universal expression system for heterologous proteins. All possess some advantages and also disadvantages that should be considered. Choosing the best expression system requires evaluating the options – from yield to glycosilation, to correct folding, to economics of scale-up [16].

2.2. Choosing an host system

Bacterial expression systems are usually the primary choice for initial screening of recombinant expression and the reason lies in the cost-effectiveness of bacteria, its well-characterized genetics, and the availability of many different bacterial expression strains. *E*. coli is used extensively as it has a fast generation time (20-60 min depending on the growth media) and there are many tools available for gene cloning and expression. Starting in chemically defined media for growing, these cells can: be grown to very high densities; accept a wide variety of plasmids with a diversity of copy number; be used with inducible promoters with various strengths; be obtained with a wide range of plasmid systems that facilitate high-throughput cloning and expression. Several specialized strains have also been developed to overcome problems observed when over expressing proteins, e.g. strains with enhanced levels of tRNAs for rare codons, strains lacking specific proteases, strains with enhanced disulfide bond formation, etc [17]. The major drawbacks of using *E*. coli for recombinant protein production are its lack of secretion systems for efficient release of proteins to the growth medium; inability to perform extensive disulfide-bond formation and other posttranslational modifications [18]; beyond this, protein expressed in large amounts often precipitate into insoluble aggregates called inclusion bodies; and initial lysis steps to recover cytoplasmatic proteins often results in the release of endotoxins, which must be removed from the final product [15].

Of all eukaryotic hosts yeast is the most used alternative system for expression of foreign proteins because it can be grown relatively rapidly (doubling time 90 min) on simple media and to high cell densities. Its genetics is better known than any other eukaryote, so it can be manipulated readily and its complete genomic sequence is available. Finally it possesses stable high copy nuclear plasmids and finally has the ability to secrete the target protein. The major disadvantage of yeast as expression systems lies in the fact that it performs glycosilation in target proteins; however both N- and O-linked oligosaccharide structures are significantly different from their mammalian counterparts [15].

Baculoviruses have also emerged as a popular system for overproducing recombinant proteins in eukaryotic cells. Several factors have contributed to their popularity: they use many of the protein modification, processing, and transport systems present in higher eukaryotic cells; they use a helper-independent virus that can be propagated to high titers in insect cells adapted for growth in suspension cultures, making it possible to obtain large amounts of recombinant proteins with relative ease; and expressed proteins are usually expressed in the proper cellular compartment, i.e membrane proteins are usually localized to the membrane, nuclear proteins to the nucleus and secreted proteins secreted into the medium. The major drawbacks in expression using baculovirus are: since baculoviruses only infect invertebrates, it is possible that the processing of proteins produced by vertebrates is different and this seems to be the case for some post-translational modifications, e.g. internal proteolytic cleavages at arginine- or lysine-rich sequences are highly inefficient; the glycosylation capability is generally limited to producing only high mannose type and not processed to complex type oligosaccharides containing fucose, galactose and sialic acid [15].

Ideally, proteins requiring mammalian post-translational modifications should be expressed in mammalian cells. If product authenticity is absolutely essential for clinical efficacy, then despite the many shortcomings, a mammalian host is the only choice, as it offers the greatest degree of product fidelity. It should, however, be noted that oligosaccharide processing is species- and cell type-dependent among mammalian cells. Differences in glycosilation pattern are reported in rodent cell lines and human tissues. Even the use of human cell line is not perfect, since the transformation event required in most cases to produce a stable cell line may itself result in altered glycosilation profiles. Also mammalian expression techniques are time consuming and much more difficult to perform on a large scale. Complex nutrient requirement and low product concentration have meant that the end product must be highly value-added for this approach to be commercially viable [15].

2.3. Choosing an expression vector

At present there are a wide range of vectors available to introduce DNA into cells: plasmids, lambda phages, cosmids, phagmids, and artificial chromosomes from bacteria, yeast or human origin (BAC, YAC and HAC, respectively) [15]. Most expression systems for *E.* coli make use of a plasmid, because it ensures easy cloning of the recombinant gene into the host cell, without need for chromosomal integration of the recombinant gene; furthermore plasmid-based systems allow many copies of the gene in a single cell. In general, expression vectors have the following attributes (**Figure 6**):

- Ori: Sequences that allow their autonomous replication within the cell.
- Promoter: A tightly regulated promoter, i.e. one which can be switched on and off easily.
- Selection marker(s): Sequences encoding a selectable marker that assures maintenance of the vector in the host.
- Terminator: A strong transcriptional terminator should be used with a strong promoter to ensure that the RNA polymerase disengages and does not continue to transcribe downstream genes.
- Polylinker: To simplify the insertion of the heterologous gene in the correct orientation within the vector [15].

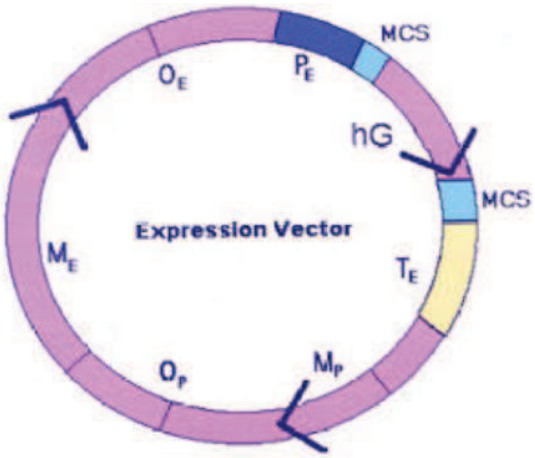

Figure 6: Design of a typical vector for heterologous gene expression MCS: multiple cloning sequence; O_p: bacterial origin of replication; M_p: marker gene for selection in

eukaryotes; O_E: eukaryotic origin of replication; P_E: eukaryotic promoter sequence; T_E: eukaryotic terminator sequence; hG: heterologous gene for expression [15].

The most popular commercial vector system for *E*. coli is the Novagen pET system based on the T7 promoter. In this system the plasmid containing the target gene is transferred to an *E.* coli strain containing a chromosomal copy of the gene for T7 RNA polymerase. There are a great variety of cell strains that can be used in order to express proteins in *E.* coli. The most widely used cell strain is BL21 which does not encode *Ion*, *ompT*, and *ompP* proteases minimizing protein degradation during purification. There are different strains within BL21. For example, BL21(DE3) contains a T7 RNA polymerase gene that is integrated into the genome and that is under control of the *lacUV5* promoter, which is a lactose analog IPTG inducible system. This means that the addiction of IPTG to a medium containing these cells will induce T7 RNA polymerase production which will ultimately transcribe the target gene in the plasmid (**Figure 7**) [19].

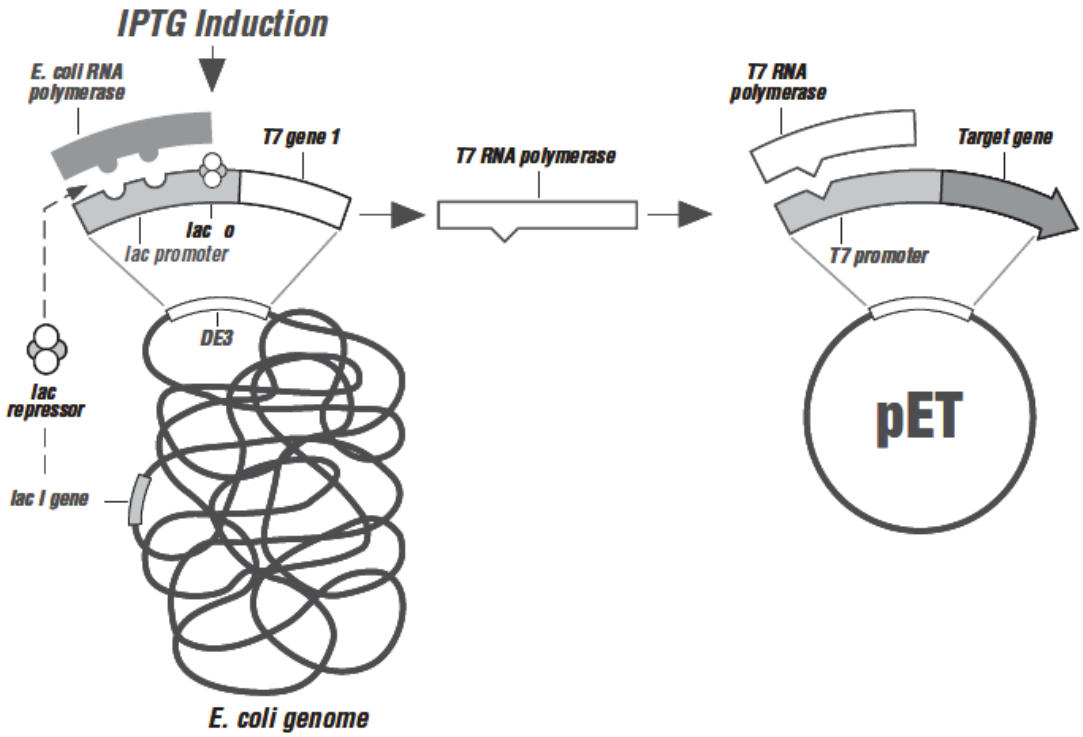

Figure 7: Induction of the *lac* gene by IPTG, in the pET system [19].

The difference between BL21(DE3) and BL21(DE3)pLysS is that the pLysS strain has a plasmid encoding a T7 lysozyme gene that allows the expression of more toxic proteins in greater amounts since that gene is a natural inhibitor of T7 RNA polymerase and therefore, protein expression is tightly regulated. Moreover, the cell wall lysis in the purification step is greatly enhanced in pLysS strain compared to DE3 because the lysozyme induces the cell lysis in addition to the chosen method of cell disruption. Other strains (**Table 1**) that can be used for protein expression are the Tuner and derivatives (Rosetta and Origami B) which are *lacY1* deletion mutants of BL21 and enable adjustable levels of protein expression throughout all cells in a culture. The *lacY1* mutation allows uniform entry of IPTG into all cells in the population, which produces a concentration-dependant, homogenous level of induction. By adjusting the concentration of IPTG, expression can be regulated from very low level expression up to the robust, fully induced expression levels commonly associated with pET vectors. Lower level expression may enhance the solubility and activity of difficult target proteins [19]. Rosetta strains are derived from BL21 *lacZY* and Tuner strains but they carry a pRARE derived plasmid that encodes several rare *E.* coli tRNA genes that can be encoded by a chloramphenicol resistant plasmid. Some of the rare codon genes are AGG/AGA (arginine), AUA (isoleucine), and CCC (proline). Therefore, these strains are designed to increase the expression of heterologous proteins whose genes encode numerous rare *E.* coli codons. Expression of these proteins can be dramatically increased when the level of rare tRNA is increased within the host. Origami strains have mutations in both the *trxB* and *gor* genes, which enhance disulfide bond formation in the cytoplasm. These strains are selectable on kanamycin, tetracycline, ampicillin, and chloramphenicol. OrigamiB is a derivative of origami and Tuner strains and it also contains *trxB* and *gor* mutations [19].

Table 1: Common applications of 5 *E*. coli strains.

General expression	Cell strain BL21	Comments T7-based expression upon IPTG induction.
Toxic Protein	BL21(DE3)	The pLysS plasmid produces T7 lysozyme to reduce basal level expression of the gene of interest.
Insoluble Protein	Tuner (DE3) pLacI	Contains a mutation in *lacZY* gene. This enables adjustment of protein expression by using different IPTG concentrations.
Low Expression	Rosetta (DE3) pLysS	They supply tRNAs rarely used in *E.* coli providing a "universal" translation.
	OrigamiB (DE3) pLysS	Contains mutations in thioredoxin reductase and glutathione reductase genes, which enhances disulfide bond formation in the cytoplasm.

2.4. Purifying an recombinant protein

2.4.1. Cell lyses – Sonication

In some biotechnological applications the biological materials (i.e. proteins, enzymes, etc) to be purified and analyzed are found inside cells and must be released by means of a process called cellular lyses. Goals of cellular lyses are: to solubilize the maximum quantity of product inside cell without promoting loss of activity and unwanted secondary modifications of the product (i.e. denaturation, proteolysis and oxidation). Lysis must be fast and effective to get as much as possible of the material from inside cell, for subsequent purification. The techniques employed in lyses must be powerful enough to destroy membrane and cell wall but at same time they must be soft to avoid changing chemically and physically the materials that are found inside cell. The technologies used can be divided in two major groups: mechanical and non-mechanical. Mechanical methods break the cell totally by application of: pressure (i.e. French Press

and High Pressure Homogenizers); mechanical agitation (i.e. Bead Mills); and ultrasound (i.e. Sonication). The non-mechanical methods induce lysis by a physical process (i.e. osmotic shock, freezing/thawing), chemical degradation (i.e. detergents, organic solvents) or by a biological process (i.e. enzymes).

2.4.2. Purification – IMAC (Immobilized Metal Affinity Chromatography)

The purification of proteins from native or recombinant sources, in their biologically active form constitutes a step of major importance in molecular biology which aims to produce a relatively large quantity of purified proteins for subsequent use. Generally, before starting a purification procedure we must be aware of some considerations about proteins. They are easily degraded and modified by microorganisms, very sensitive to temperature, pH, salts, metals, etc. Knowing all this it is of extreme importance to plan precisely the purification procedure in terms of buffer composition, temperature, sequence and duration of the steps.

Chromatographic methods are very powerful and extensively used for the purification or separation of mixtures. The reasons for this choice are the fact that they are economical and can give quantitative and qualitative information. The general configuration is of a stationary phase (matrix) packaged in a column through which the mobile phase is eluted. The solutes in mobile phase (proteins, antibodies, peptides) are adsorbed in porous medium to be later removed by means of an eluent resulting in separation of the different molecules [20]. Depending on the column proteins can be separated by means of charge (ionic change chromatography), hydrophobicity (HIC – chromatography of hydrophobic interaction), size (GF – gel filtration) or capacity to link specific chemical groups (affinity chromatography – Ni-Sepharose).

IMAC is a technique that was introduced in 1975. Is one of the most used methods for purifying polyhistidine-tagged recombinant proteins because it makes use of the natural tendency of histidine to form a complex with divalent metals around neutral pH (**Figure 8**). Basically it explores the immobilization of the metal ion on a chromatographic resin

by chelation and so this allows the separation of the histidine-tagged proteins from most untagged proteins even under denaturing conditions. The binding interaction with the tagged protein is pH dependent. However the binding can be disrupted by eluting the sample from the resin by reducing the pH and increasing the ionic strength of the buffer or by including EDTA or imidazole in the buffer (**Figure 9**) [21].

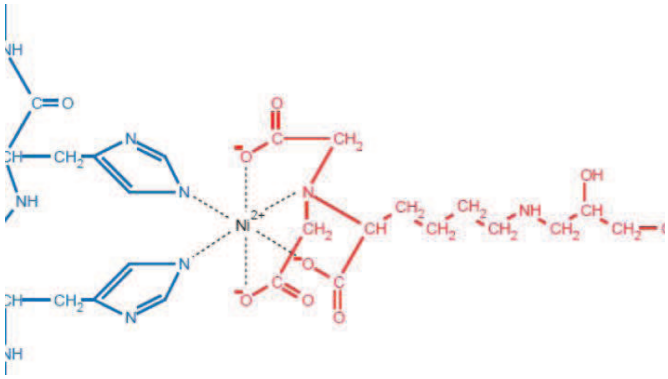

Figure 8: Interaction of histidine residues with Ni-NTA resin [22].

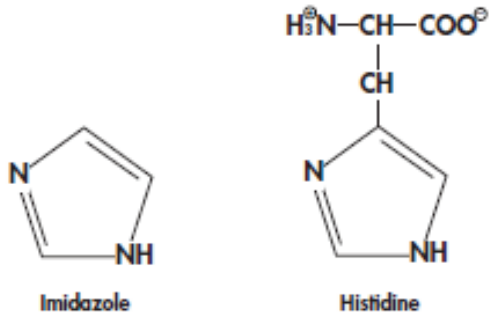

Figure 9: Imidazole molecule simple and as part of histidine residue.

2.4.3. Evaluating protein expression level: SDS-PAGE (Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis)

Electrophoresis is a technique for separation of molecules that uses the migration of molecules with charges in solution by means of the application of electric field. The rate of migration depends on the strength of the applied field, charge, size, shape, ionic strength, viscosity and temperature of the medium in which the molecules move. In non-native conditions detergents are used that will denaturate the proteins. This kind of electrophoresis is used to avoid interferences in the way that proteins are associated or folded. So denaturing and reducing agents are used to remove all levels of conformation. This method allows separation of proteins by molecular weight. It makes use of an anionic detergent, SDS (Sodium Dodecyl Sulphate) which links to proteins in an approximately 1 molecule of SDS to 3 amino acids giving them a negative charge that is proportional to their size. When denaturated proteins are introduced in an electrical field, they will move toward the positive pole (**Figure 10**). If a medium that allows proteins of different sizes move in different rates, the most used medium being polyacrilamide gel in which lower molecular weight proteins migrate at lower rates than higher ones.

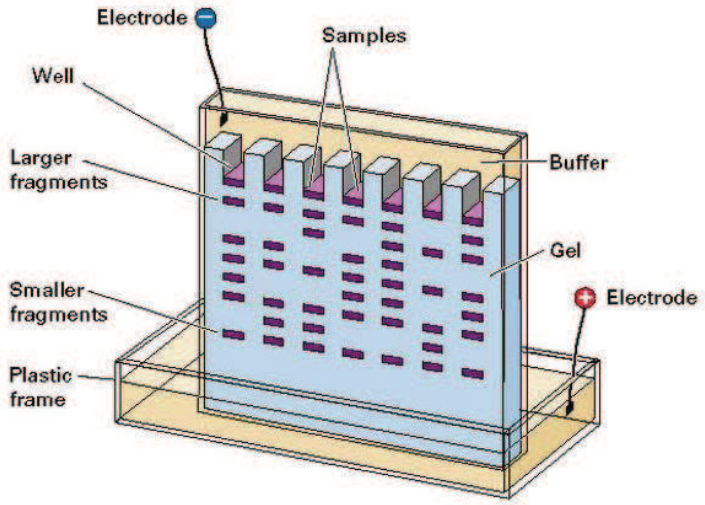

Figure 10: Electrophoresis system [23].

2.5. Structural studies

2.5.1. Protein Structure Determination

During the last decades there has been a huge growth in attempts to understand physiological processes at a molecular level. The secret of the genetic code has been discovered and we understand how this information is translated into proteins. The sequencing of whole genomes of several organisms including the human genome has also been carried out. Structural biology as a discipline started and includes the elucidation of the three dimensional structures of biological macromolecules and their interactions with other molecules in order to understand their biological function.

For understanding the functions of complex biological processes, information regarding the structure and dynamics of macromolecules is of great importance. NMR spectroscopy and X-ray crystallography are the only biophysical methods which can provide high-resolution structures of biological molecules such as proteins and nucleic acids and their complexes at atomic resolution [24]. The first protein structure determination was performed in 1959 by X-ray crystallography [25]. Since then X-ray crystallography has dominated protein structure determination for decades although recently, NMR spectroscopy has evolved into a powerful technique that complements protein crystallography [26].

When we take a look at Protein Data Bank we find that despite the growing number of proteins structures determined by NMR, X-ray crystallography is still the method that resolves most structures. X-ray crystallography is an older technique than NMR and a very robust method to obtain structural data of biomolecules but it has limitations: for crystallography we need crystals and not all proteins can be crystallized. On other hand the study of proteins by NMR is limited by the fact that not all the proteins can be solubilized in the concentrations required. Another drawback of x-ray crystallography is the fact that it determines a static molecule. The flexibility and dynamics of biomolecules are of great importance in the understanding of protein-protein and protein-ligand interactions and should be studied to obtain a complete picture. NMR is a powerful tool for structural studies because it allows us to study structural features in solution under a wide range of conditions such as temperature, pH and solvent.

Nowadays bimolecular NMR spectroscopy is a robust method used to get information about conformational dynamics and exchange processes of biomolecules at timescales ranging from picoseconds to seconds and is very efficient in determining ligand binding and mapping interaction surfaces of protein/ligand complexes [24].

2.5.2. Protein sequence-specific assignment

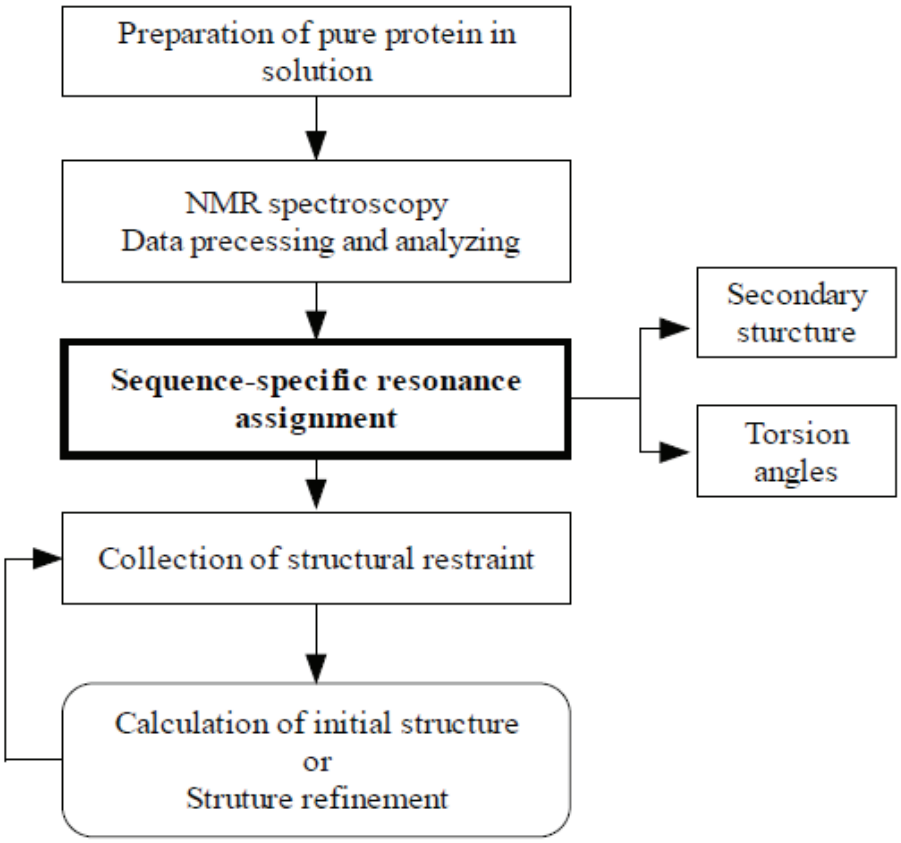

Figure 11: Flowchart of protein structure determination by NMR spectroscopy. The sequence-specific resonance assignment that is emphasized by bold plays a key role in protein structure determination [27].

Sample preparation is a very important step in structure elucidation (**Figure 11**). After producing a recombinant protein using an adequate host and expression system the last step for NMR sample purification is to choose a good buffer in which the protein is concentrated to approximately 1 mM. Phosphate buffer at pH 5–7 (20–50 mM) with or

without salt (e.g., KCl, NaCl) is often used for many NMR samples. High quality NMR tubes should be used for protein samples, which are usually tubes 5 mm in diameter containing 0.5 ml 95% H_2O/5% ^2H_2O for aqueous samples. If the volume of the sample is limited, microtubes for which the susceptibility matches that of ^2H_2O are chosen for a total sample volume of approximately 200 μl, such as Shigemi micro tubes (Shigemi Inc., Allison Park, PA). The buffer contains from 5-7% ^2H_2O used for ^2H lock. In addition, the samples are usually required to be degassed by blowing high purity argon or nitrogen gas into them to remove oxygen—the paramagnetic property of which will broaden the line shapes of protein resonances [28].

2.5.2.1. Sequential resonance assignment in unlabeled and ^{15}N labeled proteins

The application of two-dimensional NMR methods to proteins was for the first time carried out in the early 1980s by Kurt Wuthrich and it revolutionized the application of NMR for structure determination of macromolecules [29]. These methods were developed for unlabeled proteins and allowed small proteins (up to 10-12 kDa) to be assigned and distance restraints to be obtained. The first step in any structural studies of biomolecules by NMR is complete resonance assignment. For this step the methods developed by Wuthrich and co-workers are of great importance because they are based on the identification of spin systems within individual amino acids by through-bond 1H-1H connectivity measured using COSY and total TOCSY. Sequential conectivities between neighboring amino acids via NOESY are subsequently carried out. 2D NMR solved the problem of signal overlap. However the increasing number of signals made resolution to decrease and also the sensitivity due to intensity losses during the additional magnetization transfers. Therefore longer measurements times are needed.

At the end of 1980's developments in three-dimensional homonuclear experiments made it possible to overcome problems of linewidth and number of signals seen in bigger proteins. The elucidation of the structure of large biomolecules has evolved along with protein expression techniques that makes possible to obtain large amounts of

proteins and, if necessary labeled with NMR active nuclei such as ^{13}C and ^{15}N. These improvements have resulted in a substantial increase in assignment strategies because of developments in double and triple-resonance NMR experiments that makes use of uniformly $^{13}C/^{15}N$-labeled proteins. Concluding, the advances in molecular biology techniques have extended the assignment to molecular masses up to 15-20 kDa. Recently, the introduction of $^{2}H/^{13}C/^{15}N$-labeling and the TROSY technique has extended even more the limits towards 100 kDa [30]. The introduction of a third frequency dimension (^{13}C) in protein NMR allowed the signals in 2D and 3D spectra to be spread further reducing overlap (**Figure 12**). The introduction of a third dimension ^{15}N made the NMR spectra less complex than homonuclear 3D because it gave the opportunity to make assignment based on the $^{1}H,^{15}N$ correlation which is specific for every individual amino acid in the chain.

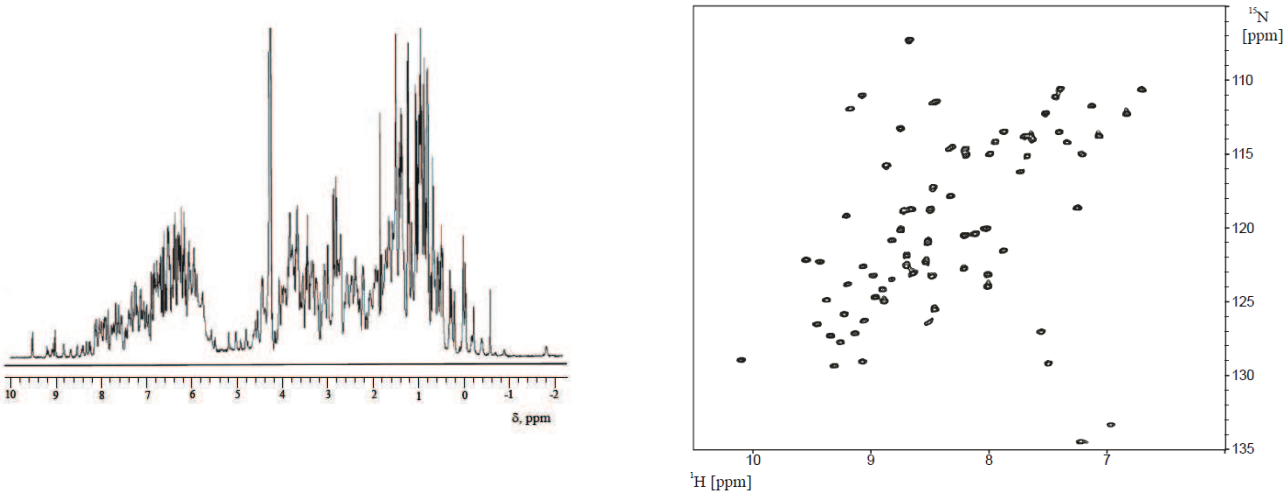

Figure 12: Protein NMR 1D (left) and 2D (right) [31].

2.5.2.2. Assignment of uniformly $^{13}C/^{15}N$-labeled proteins using triple resonance experiments

Assignments based on triple-resonance experiments make use of efficient magnetization transfer through coupled heteronuclear spins (^{1}H, ^{13}C and ^{15}N). Nowadays many three-dimensional (3D) triple-resonance NMR experiments are available to allow the sequential backbone assignment for large proteins. The names of these experiments are related to the nuclei they correlate. For example, the experiment that transfers magnetization from the amide proton (H^N) via the amide nitrogen (N) to the carbonyl C′(CO) of the previous residue is called HNCO. If we include another step of magnetization transfer to the Cα (CA) of the previous residue then the experiment is referred to as H(N)COCA, being that the parenthesis means that the magnetization is only transferred via the nitrogen spin, without chemical shift evolution (**Figure 13**).

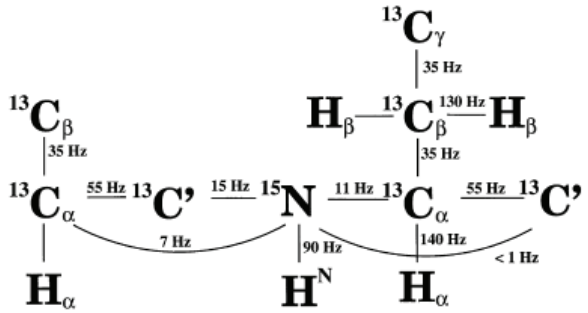

Figure 13: Spin system of the peptide backbone and size of the ^{1}J and ^{2}J coupling constants that are used for magnetization transfer in ^{13}C and ^{15}N-labelled proteins [32].

In **Figure 14**, the triple resonance experiments used during backbone spins assignment are shown. The HNCA experiment correlates the ^{1}H from the backbone amide group with the attached ^{15}N and the $^{13}C\alpha$ of the same amino acid. In addition, the $^{13}C\alpha$ of the previous residue in the protein sequence can be also detected. The HN(CO)CA experiment correlates the ^{1}H from the backbone amide group with the attached ^{15}N and the $^{13}C\alpha$ of the previous amino acid in the protein sequence. Both experiments,

HNCA/HN(CO)CA, have complementary information allowing the identification of ^{1}HN, ^{15}N and ^{13}C resonances and the attribution of sequential connectivities. The same goes for HNCO/HN(CA)CO and HN(CO)CACB/HNCACB experiments which allows the sequential assignment of the proteins under study.

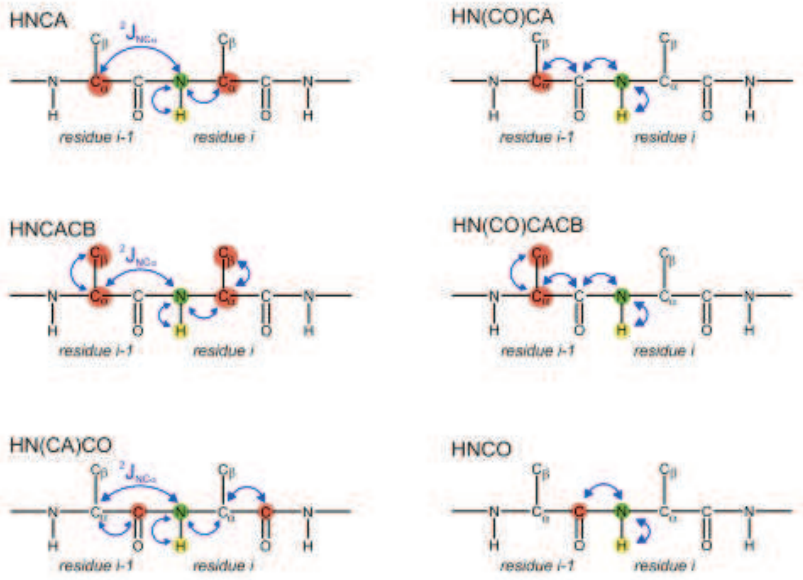

Figure 14: Heteronuclear 3D NMR experiments and correlations, respectively.

In **Table 2** we have the average chemical shifts of active nucleus (H, C and N) that are present in the 20 naturally amino acids present in proteins. This table can be really helpful upon the assignment of the resonances because some systems appear in a specific region of spectra, being directly just to look to the spectrum (i.e.^{13}Cα for glycines).

Table 2: The statistics presented in this table were calculated from the full BMRB database. This includes only the diamagnetic proteins. The calculated statistics are drived from a total of 4603403 chemical shifts in the 20 natural amino acids found in proteins (adapted from BIOLOGICAL MAGNETIC RESONANCE BANK – BMRB Database, 2011)

Residue	H^N	H^α	^{13}CO	$^{13}C\alpha$	$^{13}C\beta$	^{15}N
ALA	8.19	4.26	177.73	53.15	19	123.22
ARG	8.24	4.3	176.42	56.79	30.68	120.78
ASP	8.31	4.59	176.40	54.69	40.88	120.65
ASN	8.34	4.67	175.27	53.54	38.69	118.93
CYS	8.39	4.66	174.87	58.24	32.66	120.13
GLU	8.33	4.25	176.89	57.35	30	120.66
GLN	8.22	4.27	176.32	56.59	29.18	119.88
GLY	8.33	3.94	173.88	45.36	-	109.65
HIS	8.25	4.61	175.25	56.49	30.24	119.66
ILE	8.27	4.18	175.85	61.63	38.61	121.45
LEU	8.22	4.32	176.99	55.64	42.30	121.83
LYS	8.18	4.26	176.65	56.96	32.77	121.04
MET	8.26	4.41	176.2	56.12	32.99	120.09
PHE	8.36	4.63	175.43	58.11	39.95	120.47
PRO	-	4.4	176.73	63.34	31.85	133.96
SER	8.28	4.48	174.64	58.74	63.79	116.26
THR	8.24	4.46	174.57	62.23	69.72	115.41
TRP	8.29	4.68	176.13	57.68	29.98	121.67
TYR	8.32	4.63	175.4	58.13	39.32	120.52
VAL	8.29	4.18	175.63	62.51	32.72	121.12

2.5.3. TROSY (Transverse Optimized Relaxation Spectroscopy) in the study of large proteins

The foundations of NMR resonance assignment studies are high-quality NMR spectra recorded with good S/N ratio and spectral resolution. However, with increasing molecular weight these requirements become harder to achieve. The limiting factors are sensitivity and line broadening due to rapid transverse spin relaxation and extensive signal overlap due to the high complexity of the spectra. Recent advances have been achieved with both novel NMR techniques and new biochemical approaches. In particular, using the NMR technique Transverse Relaxation-Optimized Spectroscopy (TROSY) in combination with suitable isotope labeling schemes, the size limit for the observation of NMR signals in solution has been extended severalfold [33-34].

The TROSY (transverse relaxation-optimized spectroscopy) technique has been developed to reduce relaxation losses during the chemical shift evolution of a heteronucleus X (e. g. ^{15}N), the X→ ^{1}H magnetization transfer and the acquisition time. The transverse relaxation rate is mainly caused by DD coupling and CSA. Considering the components of a doublet of I in a weakly coupled two spin system IS the transverse relaxation rates are different due to addition or subtraction of the influence of DD coupling and the CSA. A narrow and a broad component is the result. In a non-decoupled two-dimensional ^{15}N-HSQC spectrum only one component of the quartet has a narrow line width in both dimensions (the TROSY component) as shown in below. TROSY experiments select solely these narrow components and suppress the broad components of the quartets (**Figure 15**). Note that the signals in a TROSY experiment are shifted in both dimensions by $^{1/2}J_{NH}$. To summarize, TROSY suppresses transverse relaxation, e. g. in ^{15}N-^{1}H moieties by constructive use of interference between DD coupling and CSA [31].

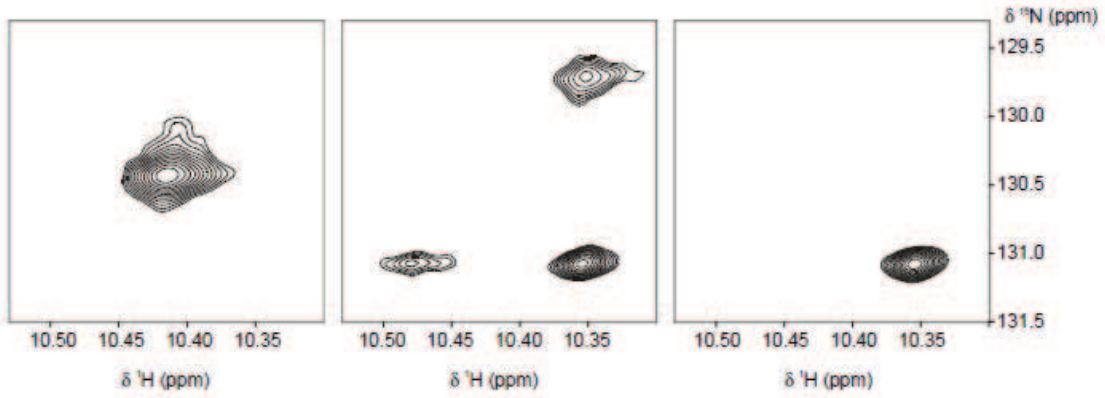

Figure 15: (a) conventional 15N-HSQC with decoupling in both dimensions, (b) 15N-HSQC without decoupling in both dimensions and (c) 15N TROSY which selects only the narrow component [31]**.**

The TROSY technique can be implemented in every NH detected triple-resonance experiment and is widely used combined with triple resonance experiments. The advantages are higher resolution of the experiments and smaller half height line widths. Therefore the separation of signals in the $^{15}N,^{1}H$ plane is better [35].

2.6. Interaction studies

Generally, the biological function of a protein is intimately related to its interaction with ligands. Examples are the interaction of hormone receptors with hormones, that trigger complicated signal cascades, or the interaction of certain proteins with nucleic acid sequences to regulate gene replication, transcription, or translation, or the highly specific interaction of cell-surface antigens with receptors located on other cells [36]. Protein-ligand interactions are an extremely important field of study. Information on protein-ligand interactions not only plays an integral role in the fundamental investigation of any enzyme, cellular receptor, or other large biomolecules and the study of its binding properties, but is also important because of its utility in the discovery and development of new drugs. Nowadays there is a large list of techniques used to access this kind of information, but in this thesis we will focus in two of them, fluorescence quenching and chemical shift mapping.

2.6.1. Chemical shift mapping

NMR can be used to solve 3D structures in solution; however it is also a very powerful tool for accessing interactions between macromolecules and various ligands. This interactions can be assessed at a wide variety of levels, being the chemical shift mapping being one of the possible approaches [37]. In theory, upon binding phenomenon of a ligand to a receptor protein, the chemical shifts of both ligand and protein proton resonance signals will be affected. For the protein, the most affected resonances are from the nuclei located in the binding pocket (**Figure 16**). The acquisition of $^{1}H/^{15}N$ HSQC spectra and assignment of the spin systems with and without ligand present is, generally one of the possible experimental setups to detect ligand binding and at same time identify the amino acids present in the binding pocket. The technique only requires that ^{15}N-labeled protein is available [36].

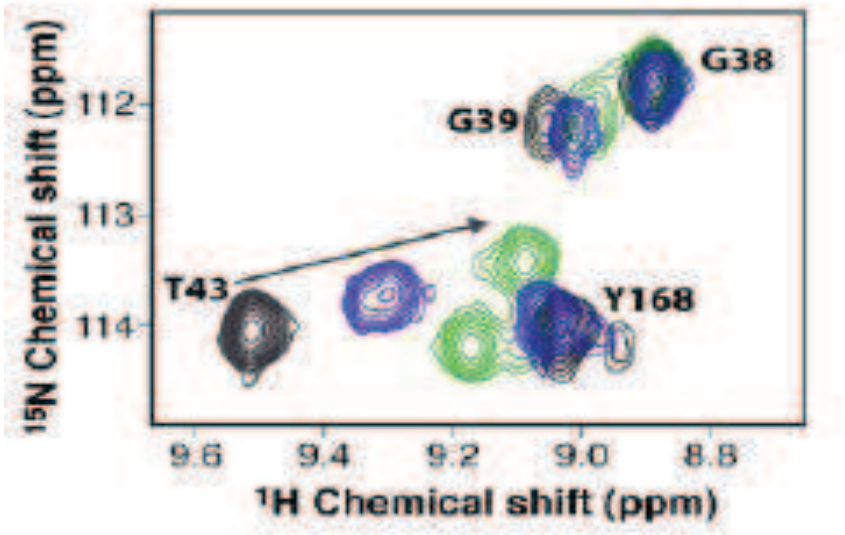

Figure 16: TROSY of chemical shift perturbation upon binding of murine p22HBP (black) to PPIX (blue) and hemin green) [9].

2.6.2. Fluorescence quenching

The interactions of proteins with other molecules are of great importance because they are responsible for the majority of molecular complexes that makes life possible. So it is of great importance to quantify the number and binding strength of protein-ligand interactions. Fluorescence techniques are largely used to study these kinds of interactions because of the sensitivity of the technique which allows it to be used in real equilibrium situations [38]. Luminescence is the emission of light from any substance and it takes place from electronically excited states. Luminescence can be divided into fluorescence and phosphorescence depending on the nature of the excited state. In the singlet excited state, an electron in the excited orbital is paired (opposed spin) with an electron in the ground state. Here the return to the ground state of the excited electron is allowed and occurs rapidly (fluorescence). Phosphorescence is the emission of light from triplet excited states, in which the electron in the excited orbital has the same spin orientation as the electron in ground state. Here the transitions from excited to ground are not allowed the return to ground state is much slower than in fluorescence (**Figure 17**) [39].

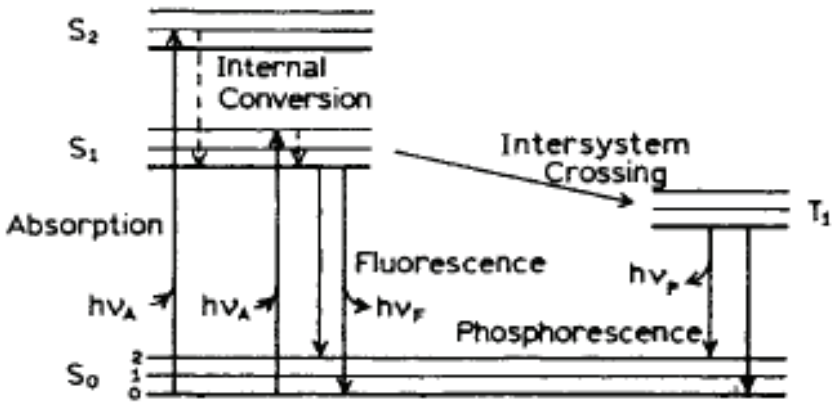

Figure 17: Jablonsky diagram showing the light absorption and emission phenomenon [39].

Fluorescence quenching refers to any process that diminishes the sample fluorescence intensity. A large range of molecular interactions can result in quenching. These include excited state reactions, molecular rearrangements, energy transfer and collisional

quenching. Fluorescence quenching has been largely studied both as fundamental phenomenon and as source of information for biochemical systems. Biochemical applications are related to the role of intrinsic interactions in quenching. Both static and dynamic quenching requires molecular contact between fluorophore and quencher. In collisional quenching, the quencher has to collide with the fluorophore for the excited state lifetime. Due to this contact, the fluorophore returns to the fundamental state without photon emission. Generally quenching occurs without any definitive change in the molecules or without any photochemical reaction. In the static quenching situation a complex is formed between fluorophore and the quencher and this complex is not fluorescent. The discussion of biochemical fluorescence generally concentrates on proteins. This is due to the fact that only proteins possess a useful intrinsic fluorescence. Lipids, membranes and saccharides do not present fluorescence and the intrinsic DNA fluorescence is too weak to extract important information. In proteins, the aromatic amino acids phenylalanine, tyrosine and tryptophan present fluorescence. A great advantage of proteic structure is the fact that these three amino acids are relatively rare in proteins. Of the three, the most abundant is tryptophan but its only present in 1% (mol) in proteins. An important characteristic of protein intrinsic fluorescence is the high sensitivity of tryptophan to its local environment. We can watch changes in tryptophan emission spectra in response to protein conformational changes, subunit association, substrate binding or denaturation, as the local environment which surrounds indol ring is modified (**Figure 18**) [39].

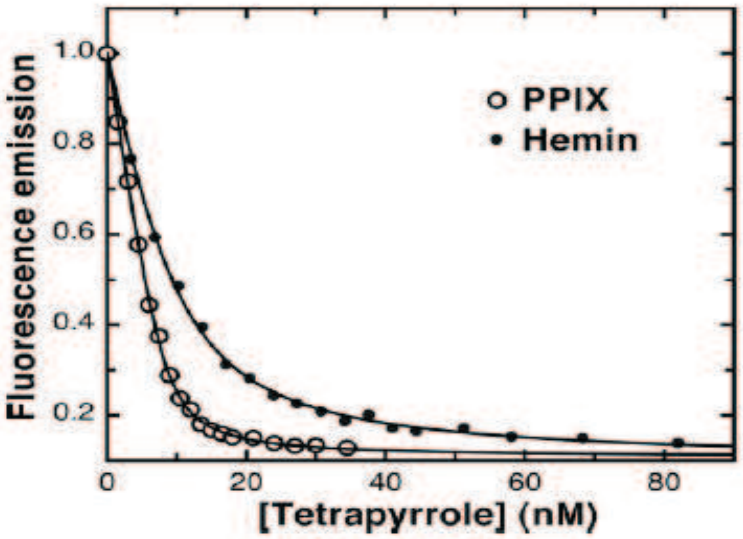

Figure 18: Tryptophan intrinsic fluorescence in murine p22HBP upon titration with PPIX and hemin [9].

Chapter 3 – Materials and Methods

3.1. Human p22HBP expression system construction

The gene encoding N-terminal domain of *Homo sapiens* recombinant p22HBP (GenBank accession number NP_057071) (**Figure 19**), consisting of aminoacids 1 to 189 (**Figure 20**), was subcloned in Nco 1 and Xho 1 sites into similarly restricted pET28a expression vector (Novagen), after the T7 promoter and the His-Tag (**Figure 21**). This plasmid was used to transform *E.* coli strain BL21 (DE3), a step carried out by Nzytech (genes & enzymes, Ltd). Long-term storage of hosts is best carried out as glycerol stocks. For this purpose after growing a culture containing kanamycin until the OD600 reached 0.6-0.8 we removed 0.375 mL and transferred them to an autoclavated eppendorf tube and added 0.125 mL 80% glycerol. After mixing well the eppendorfs were stored at -70ºC. To inoculate a culture from the frozen stocks a few microliters at the thawed stock is pippeted to fresh medium.

```
CCATGGGCCATCACCATCACCATCACCTGGAACTGGGTATGATTAAAAACAGTCTGTTTGGCAGC
GTGGAAACCTGGCCGTGGCAGGTTCTGAGCAAAGGTGATAAAGAAGAAGTGGCCTACGAAGAACG
TGCATGCGAAGGCGGTAAATTCGCAACCGTGGAAGTTACGGATAAACCGGTGGATGAAGCACTGC
GCGAAGCGATGCCGAAAGTTGCCAAATATGCAGGCGGTACCAATGATAAAGGCATTGGTATGGGC
ATGACGGTGCCGATCTCTTTTGCGGTTTTCCCGAACGAAGATGGTAGTCTGCAGAAAAAACTGAA
AGTGTGGTTTCGTATCCCGAATCAGTTTCAGAGCGATCCGCCGGCACCGAGCGATAAATCTGTGA
AAATTGAAGAACGCGAAGGCATCACCGTTTACTCTATGCAGTTTGGCGGTTACGCGAAAGAAGCC
GATTATGTTGCGCAGGCCACCCGTCTGCGTGCCGCACTGGAAGGTACCGCCACGTATCGTGGCGA
TATTTACTTCTGTACCGGTTATGATCCGCCGATGAAACCGTATGGCCGTCGCAACGAAATCTGGC
TGCTGAAAACGTAACTCGAG
```

Figure 19: Optimized sequence of the gene encoding recombinant *Homo sapiens* p22HBP. Start and Stop codons underlined.

10	20	30	40	50	60
MGHHHHHHLE	LGMIKNSLFG	SVETWPWQVL	SKGDKEEVAY	EERACEGGKF	ATVEVTDKPV
70	**80**	**90**	**100**	**110**	**120**
DEALREAMPK	VAKYAGGTND	KGIGMGMTVP	ISFAVFPNED	GSLQKKLKVW	FRIPNQFQSD
130	**140**	**150**	**160**	**170**	**180**
PPAPSDKSVK	IEEREGITVY	SMQFGGYAKE	ADYVAQATRL	RAALEGTATY	RGDIYFCTGY
190					
DPPMKPYGRR	NEIWLLKT				

Figure 20: *Homo sapiens* p22HBP protein sequence. RED: His-Tag.

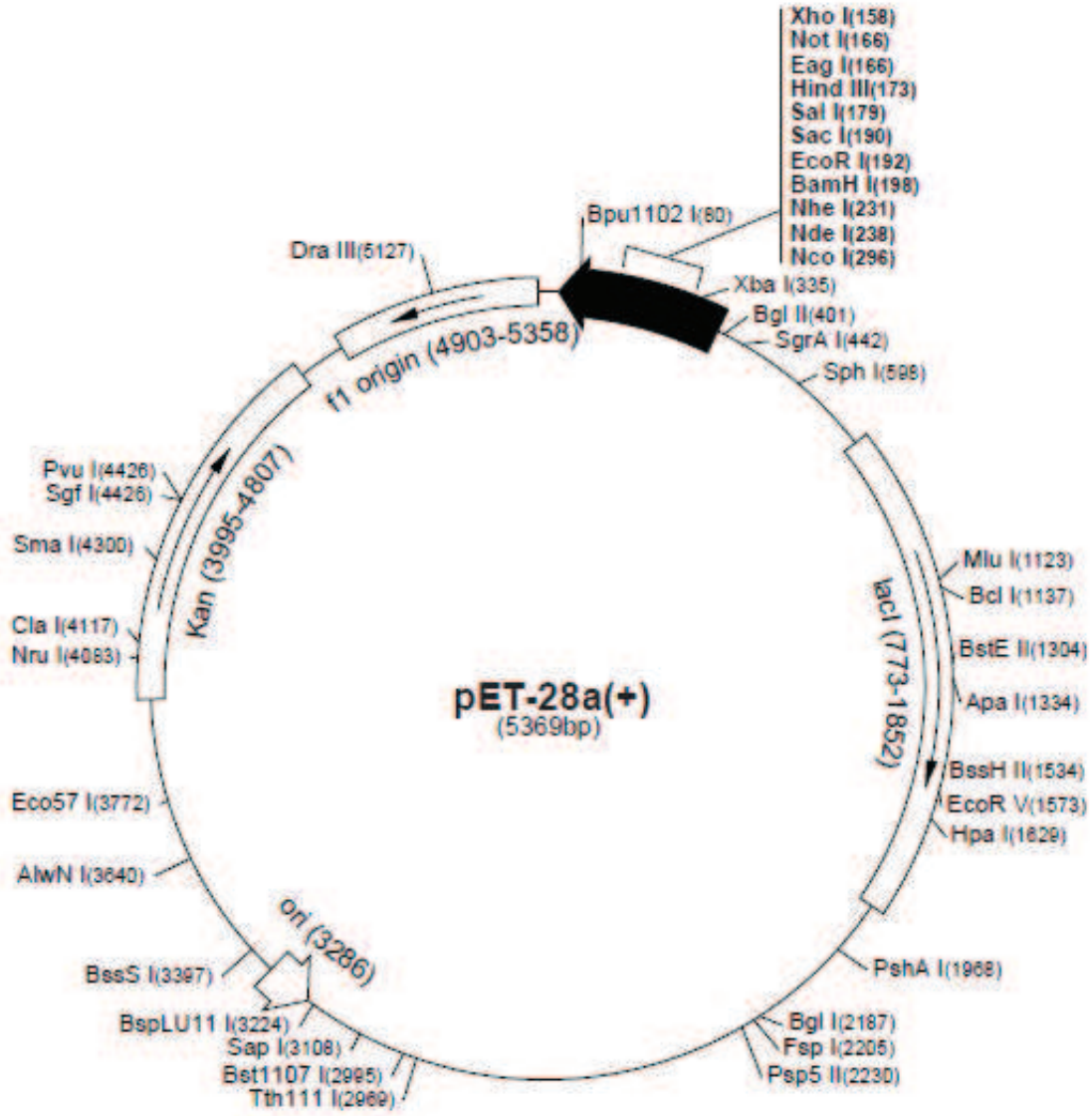

Figure 21: pET-28a map.

3.2. Human p22HBP overexpression, lysis and purification

3.2.1. Overexpression

Luria Bertani medium was used for the growth of the BL21(DE3) strain and M9 Medium was used for induction of recombinant protein expression.

LB Medium	Tryptone, yeast extract, NaCl; deionized water to 1 liter (pH 7.5). Sterilize by autoclaving.
M9 Medium (Minimal)	50 ml 10xM9 salts* plus 950 mL deionized H_2O. Sterilize by autoclaving.
***10xM9 salts**	128 g $Na_2HPO_4.12\ H_2O$; 30 g KH_2PO_4; 5 g NaCl; 10 g $(NH_4)Cl$. For ^{15}N labeled protein production, use 1 g $^{15}(NH_4)Cl$ (CORTECNET, France) moments before cell inoculation.

The cells were grown using a standard procedure in which LB medium is used to achieve good cell density and a M9 minimum medium is subsequently use for induction (**Figure 22**). For overexpression purposes, starter cultures were prepared by inoculating 20 μL of cells taken from glycerol stocks in 20 mL of LB media previously autoclaved and supplemented with kanamycin, and incubated at 37ºC with shaking at 180 rpm for 12 to 16 hours (overnight). The culture was then inoculated into LB medium, 3 mL of starter culture for 150 mL of LB, and incubated for 5 hours at 37 ºC and 180 rpm. The cells were then harvested by centrifugation at 8000 rpm for 10 min at 4ºC and ressuspended into a freshly prepared M9 medium supplemented with kanamycin and nutrients. For testing the time of induction, 4 flasks were used. In one flask the cells were induced at the time of ressuspension in M9 medium; in the second the cells were induced after 1h in M9 medium at 30ºC and 180rpm; in third and fourth the induction was performed after 2 and 4h respectively. For testing inducer concentration, 4 flasks were used and to each was added a different stock concentration of 150 μL of IPTG (0, 0.1, 0.5 and 1M) after 2h in M9 medium at 30ºC and 180 rpm

(**Table 3**). After induction, all cells were incubated at 30ºC and 180rpm for 16h (overnight).

Table 3: Parameters tested for the human p22HBP over-expression optimization.

Time of induction (hours)	IPTG concentration (mM) ± 0.040
0	0
1	0.1
2	0.5
4	1

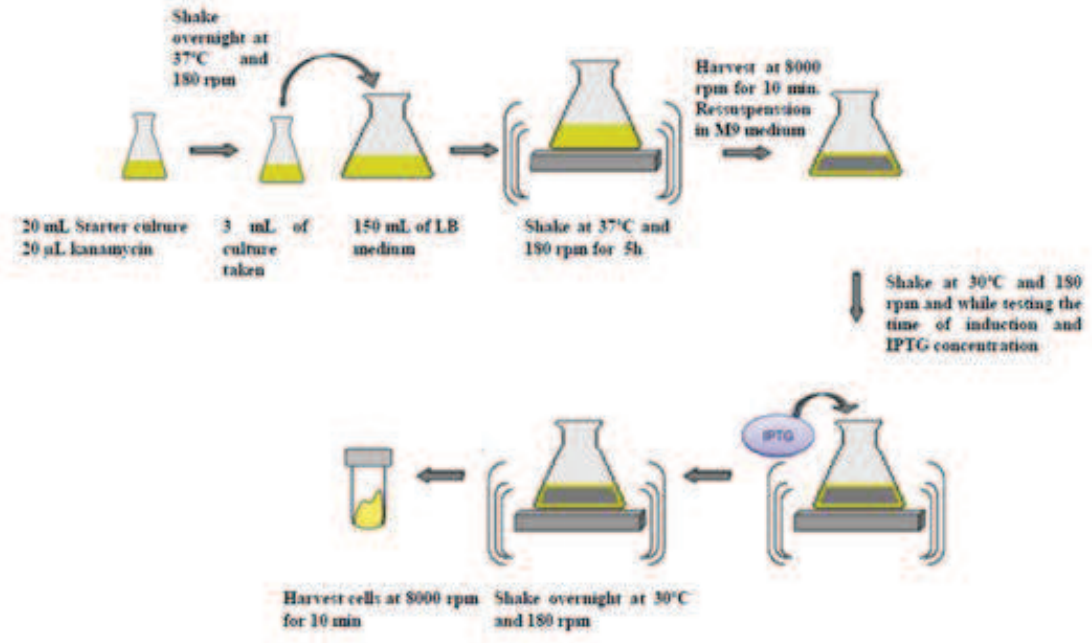

Figure 22: Sequential steps of overexpression of the human p22HBP.

3.2.2. Purification

After 16h of induction in M9 medium, the cells were recovered by centrifugation at 8000 rpm for 10 min and resuspended in a 50 mM KH_2PO_4/300 mM NaCl buffer at pH 8.0. For optimization purposes cells lysates were sonicated for 6, 12 and 20 minutes (59 seconds with 59 second intervals) while the suspension was kept on ice to prevent heating. Centrifugation at 20000 rpm at 4ºC for 1h was applied to separate the homogenate components into cytoplasmatic fraction (supernatant) and cell pellet. The supernatant was then applied to a Ni-NTA column with Sepharose matrix functionalized with chelating groups charged with Ni^{2+} for purification. The system was equilibrated with 50 mM KH_2PO_4/300 mM NaCl buffer and the purification procedure was performed under native conditions. The hist-tagged protein was eluted using an imidazole gradient (10-500 mM) (**Figure 23**). The purest fractions containing human p22HBP were pooled together and concentrated using 10 kDa centricons (Millipore) to 2.5 mL. This volume was applied to a PD-10 Desalting column (Amersham Biosciences), previously equilibrated with 20 mL of 50 mM phosphate buffer and eluted with this buffer to a final volume of 3.5 mL.

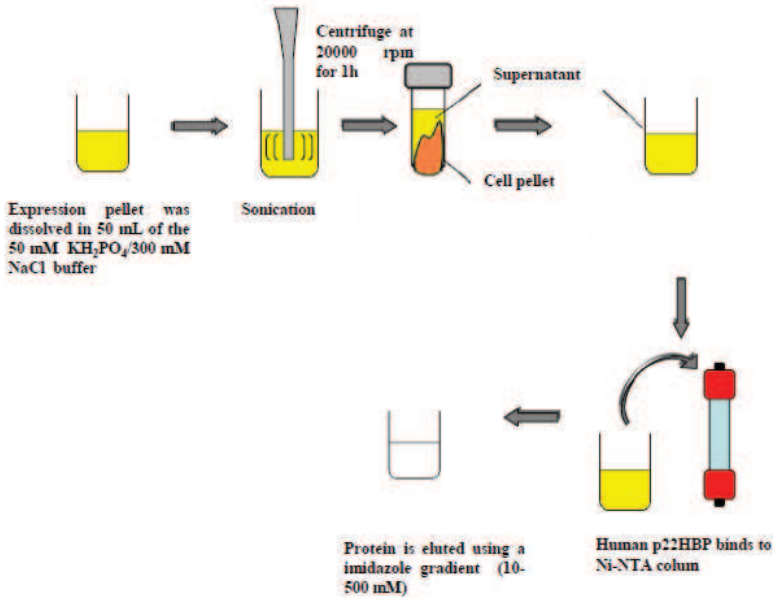

Figure 23: Sequential steps of purification of the human p22HBP.

3.2.3. Denaturing gel electrophoresis

SDS-PAGE was used to examine purity of the protein separation as well as to determine its molecular weight. A Vertical Mini-Protean 3 unit (BioRad) was assembled for SDS-PAGE gel running with a 15% running gel and 4% stacking gel prepared prior to electrophoresis. The separating gel was prepared by adding 1.7 ml of H_2O, 2.5 mL Tris-HCl 1.5 M at pH8.8, 3.7 mL of 40% acrylamide, 2.01 mL of 2% Bis-acrylamide, 100 uL of 10% SDS, 50 uL of 10% APS and 5 uL of 5µl of TEMED. A layer of water was gently loaded onto the top of the separating gel to flatten the gel surfaces and remove trapped bubbles. Subsequently, the water was removed from the gel cassette. The stacking gel, composed of 2.9 mL H_2O, 1.25 mL Tris-HCl 1.5 M at pH6.8, 500 µL of 40% acrylamide, 250 µL of 2% Bis-acrylamide, 50 µL of 10% SDS, 25 µL of 10% APS and 5 uL of 5µl of TEMED, was pipetted on the separating gel. A comb was inserted into the stacking gel before polymerization occurred. Each of the samples was mixed with 125 µl of Tris-HCl 1.5 M at pH 6.8, 75 µL glycerol, 200 µL of 2% SDS, 100 µL of 2-mercaptoethanol and bromophenol blue, followed by subsequent denaturing step at heating at 95°C for 5 minutes. Next, the samples together with the SDS marker were loaded into the respective wells of the SDS-PAGE gel. The gel was covered then with water and left to boil in a microwave oven, followed by 5 min of cooling at room temperature with slow shaking. The water was then removed and fresh water added to the gel and the procedure repeated 2 times. For the last cycle, the gel was stained with Coomassie Blue before the microwave step (containing 70 mg of Coomassie Blue G250). After removing the Coomassie solution, water was added to the gel and shaking for 2 min.

3.2.4. UV-Visible Spectroscopy

UV-Visible absorption measurements (at selected wavelengths or as 190-600 nm spectra) were recorded on a UV-visible spectrophotometer interfaced with a personal computer using the Chemstation Software for Windows. Optical measurement was performed at room temperature. Purified protein samples were quantified by UV-Vis spectrophotometric method, according to the Lambert-Beer law, by measuring the absorption at 280 nm (A280) (**Equation 1**).

$$A(\lambda) = \varepsilon(\lambda).c.d$$

Equation 1: Equation used for the determination of the human p22HBP concentration. Where A(λ) and ε(λ) are the absorption and molar absortivity at 280 nm, c is the protein concentration in M and d is the thickness of the cell (1 cm).

A theoretical value for molar absortivity of the human p22HBP was calculated by the Expasy protparameter tool and gave 35535 $M^{-1}cm^{-1}$. For calculation of the real (experimental) value to be use, a sample of 37 µM of the protein was diluted 1:2, 1:3, 1:4, 1:5, 1:10 and 1:20 and a graph of absorbance as a function of the concentration was plotted (**Figure 24**). The experimental value for human p22HBP was found to be 35686 $M^{-1}cm^{-1}$.

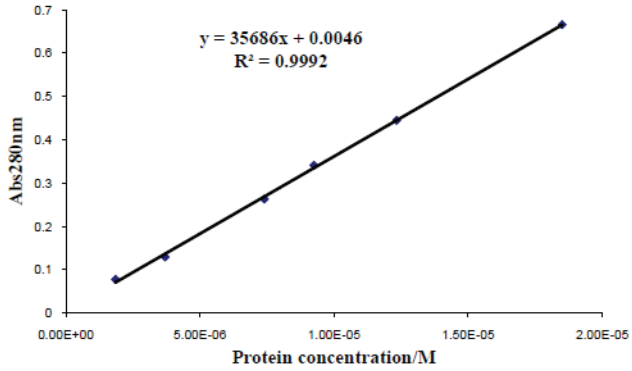

Figure 24: Plot of Absorbance values at 280 nm as a function of protein concentration.

3.3. Mass Spectrometry

A mass spectrum of 1.0 nM human p22HBP (prepared from dilution of a sample containing 20 μM of protein), was obtained on a matrix assisted laser desorption/ionization–time of flight MALDI TOF/TOF mass spectrometer (4800 Proteomics Analyzer, Applied Biosystems, Foster City, CA, USA) in the linear mid mass positive ion mode and obtained in the mass range from 5000-80000 Da with 1650 laser shots, using alpha-cyano-4-hydroxycinnamic acid as a matrix.

3.4. Fluorescence Quenching

The p22HBP samples used in the fluorescence quenching experiments were diluted from a 0.2 mM stock solution with 50 mM KH_2PO_4 at pH 8.0 and the protein concentration of the stock solution was estimated by UV spectroscopy using a molar extinction coefficient at 280 nm (ε_{280}) of 35686 $M^{-1}cm^{-1}$. Hemin and PPIX solutions were dissolved in ammonia followed by dilution in water. For the fluorescence quenching experiments 1 mg of hemin and PPIX were dissolved in ammonia followed by dilution in water. The surfactant Tween 80 was added to a final concentration of 1.5% (v/v) and the pH was adjusted to 8.0 by adding 1M NaH_2PO_4. The final concentrations prepared for both tetrapyrrole were 5 and 10 μM. In the fluorescence quenching experiments increasing volumes of the porphyrins were added to 2 mL of protein at 100 nM and the fluorescence intensities were registered (**Table A1** (Appendix)). After each addition the reaction was gently homogenized and left to equilibrate for 1 minute. Each titration experiment was repeated two times and with different preparations of porphyrin dilutions. Determination of the K_D values were performed by plotting the human p22 HBP emission maxima as a function of porphyrin concentration. In addition a fitting (OriginPro software) was made with the data obtainted by using **Equation 1** [9].

$$y = I_0 - (I_0 - I_{int}) \frac{\sqrt{K_D - 3.[HBP].x - (K_D - [HBP].x)^2}}{2.[HBP]}$$

Equation 2: p22 HBP emission maxima (y) as a function of porphyrin concentration (x). Where I_0 and I_{int} are emission intensities at zero and saturating porphyrin concentrations, respectively and [HBP] is the protein concentration.

3.5. Nuclear Magnetic Resonance Spectroscopy

3.5.1. Protein overexpression and purification

The bacterial growth was performed in 2 L LB medium, at 37ºC and 180 rpm, until an OD600 of about 0.6-0.8 was obtained. The cells were harvested by centrifugation at 8000 rpm for 10 min and resuspended in M9 medium (minimal growth medium) where ^{15}N-labeled NH_4Cl and/or ^{13}C-glucose was used as sole carbon source. The cells were let to grow for 2 h at 30ºC and 180 rpm. After then the protein induction was then started with the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) 0.5 mM and continued overnight at 30ºC. Protein purification was achieved as described in section 4.2.2.

3.5.2. Nuclear Magnetic Resonance spectroscopy experiments

For the sequential assignment of the human p22HBP, a uniformly $^{15}N/^{13}C$-labeled sample (50 mM phosphate buffer, pH 8.0) in H_2O/D_2O (7%) at a concentration of 1 mM was used. All NMR spectra (**Table 4**) were acquired at 293 K on a Bruker DRX600 spectrometer equipped with a four-channel probe and triple axis gradients. The assignments of ^{1}H, ^{15}N, ^{13}CO and $^{13}C\alpha$ chemical shifts were determined by a series of TROSY-type triple resonance experiments: HNCO, HNCA and HN(CO)CA. All the experiments were processed and analyzed by NMRPipe/NMRDraw (DELAGLIO *et al.)* and CARA (DAMBERGER *et al*.).

Table 4: Table listing the NMR experiments performed or structural studies.

Experiment	Base frequency ^{1}H (MHz)	^{1}H sweep width (Hz)	^{13}C sweep width (Hz)	^{15}N sweep width (Hz)	Acquisition time
TROSY	600	8417.509		2432.713	2h 47min
tr-HNCA		10775.863	4829.15	2189.445	1day1 4h 40min
trHN(CO)CA		10775.863	4829.15	2189.445	2days 2h 27min
trHNCO		10775.862	3320.46	2189.445	1day 3h 21min

3.5.3. Chemical shit perturbation mapping

The porphyrins stock solutions were prepared as previously described for the fluorescence quenching experiments. Titration experiments were performed with uniformly ^{15}N-labeled human p22HBP using three samples: human p22HBP:hemin and human p22HBP:PPIX (0.5 mM protein, 50 mM KH_2PO_4 buffer, pH 8.0, 7% $^{2}H_2O$). A blank solution (where no porphyrin was added) was also prepared to match the porphyrins solution conditions (50 mM KH_2PO_4 at pH 8.0 with Tween 80 at 1.5% (v/v)). The p22 HBP NMR samples were prepared by dilution of a 0.2 mM ^{15}N p22 HBP stock solution to 1 mM in 50 mM KH_2PO_4 pH 8.0 by adding 1.1 moles of porphyrin solution. The final volume was completed with the blank solution so that all the samples were under the same buffer conditions.

The spectra were acquired at 293 K, on Bruker spectrometer (500 MHz ^{1}H frequency) triple axis gradients. ^{1}H, ^{15}N TROSY spectra were acquired using a spectral width of 8012.82 Hz for ^{1}H and 2027.351 Hz for ^{15}N. The NMR data were processed as referred in section 4.5.2.

Chapter 4 – Results and Discussion

4.1. Optimizing human p22HBP overexpression, lyses and purification

In order to determine the optimum starting time for induction, four shake-flask fermentations were performed in parallel, and each culture was induced at different times of growth. In **Figure 25** the growth behavior of recombinant *E.* coli BL(21) cells containing the overproducing plasmid pET-28a is shown. A lag phase of one hour was observed in the growth of the cells. The exponential phase was completed within 10 hours of inoculation, after which the cells entered their stationary phase. In the right side we have the SDS-PAGE image of human p22HBP purification by Ni-NTA column. The image shows a good expression of the protein in different conditions and the position of the bands suggests that the protein purified has more than 20 kDa which is similar to that predicted for human p22HBP. Only with the gel image we cannot say where the maximum expression was achieved. For that, we have concentrated and quantified the purest fractions (75, 175 and 500 mM imidazole) for each induction moment and the results are listed in **Table 5**.

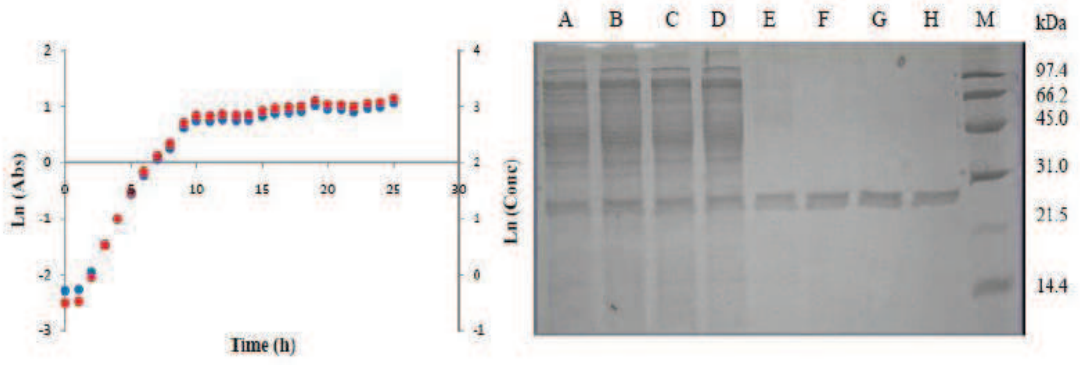

Figure 25: LEFT: Growth curve of human p22HBP. In the vertical axes we have the Logarithm of Absorbance (BLUE) and biomass concentration (RED) as a function of the number of hours. **RIGHT:** SDS-PAGE image of human p22HBP testing of the induction moment. A-D: 0 to 4h supernatant; E-H: 0 to 4h 75 mM imidazole; M: molecular weight standard.

In the **Table 5** human p22HBP concentrations resulted from time of induction testing were registered and the analysis of the values shows a direct relation between the time of induction and the amount of protein over-expressed and purified up to 2h. We can see that protein concentration increases with the time. However the data also shows that when we induce the cells after 4h of incubation in M9 medium the protein concentration decreases. Analyzing the growth curve (**Figure 25**), we can see that during the first two hours the cells are in the lag phase, where in theory most of the cellular machinery is directed to keep a regular number of cells during an adaptation period in a new medium (M9). After 2 hours in M9 medium, we can see that the cells are in an exponential growth phase. In this phase the cell machinery is directed to the synthesis of the target gene while growing. When inducing protein synthesis after 4 hours of adaptation in M9 medium, the data shows that the amount of protein produced is lower than that verified after 2 hours, even though the cells are still in an exponential growth phase where the heterologous product production should increase as the cell concentration increases. The decreasing amount of protein produced was expected only when cells start reaching the stationary phase, but this situation was not the verified. The results listed below (**Table 5**) suggests that the best time to induce the production of human p22HBP by *E.* coli BL21(DE3) containing pET-28a plasmid, is after 2h of growing the cells in M9 medium.

Table 5: Relation between the time of induction and the human p22HBP purified from the purest fractions (75, 175 and 500 mM imidazole)

Time of Induction (h)	Concentration (µM)* ± 0.040
0	14.2
1	16.2
2	19.2
4	17.5

*protein volume = 3.5 mL

After determining the optimum starting time of induction, the effect of the inducer concentration on the efficiency of the induction was examined by using 0, 0.1, 0.5 and 1 mM IPTG. The cells were all induced at the same time, after 2 hours of growth in M9

medium. The gel image of important fractions and the protein concentrations are shown below (**Figure 26** and **Table 6**).

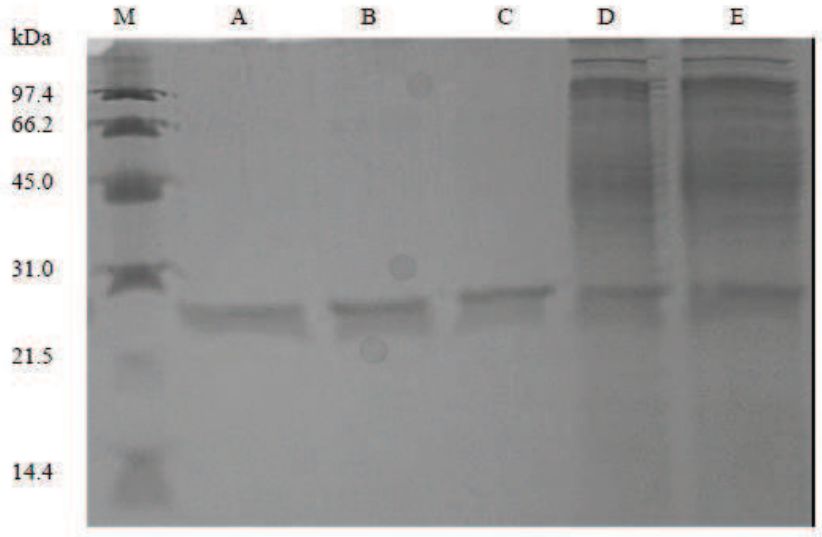

Figure 26: SDS_PAGE image of IPTG concentration testing. M: molecular weight standard; A-C: 0.1, 0.5 and 1.0 mM IPTG from 75 mM imidazole; D-E: 0.5 and 1.0 mM IPTG from supernatant.

Table 6: Relation between the concentration of IPTG used and the amount of protein produced and purified.

IPTG concentration (mM)	Concentration (µM)* ± 0.040
0	0.0
0.1	22.5
0.5	20.8
1	21.4

The SDS-PAGE (**Figure 26**) shows protein bands at positions that were predicted for human p22HBP. The comparation of protein concentrations (Table 6) shows, as expected, that IPTG concentration is an important factor in T7 RNA polymerase production, which is fundamental for transcription of the target genes. This is concordant with the absence of protein produced when not inducing the cells with any IPTG. The analysis of all results suggests that 0.1 mM of IPTG is the best concentration to use, which gives good protein yields using a minimum amount of the inducer.

Cell lysis was performed by sonication and also optimized. The parameter tested was the sonication time. Here three different periods of Sonication were tested: 6, 12 and 20 min. The results are listed below (**Figure 27** and **Table 7**).

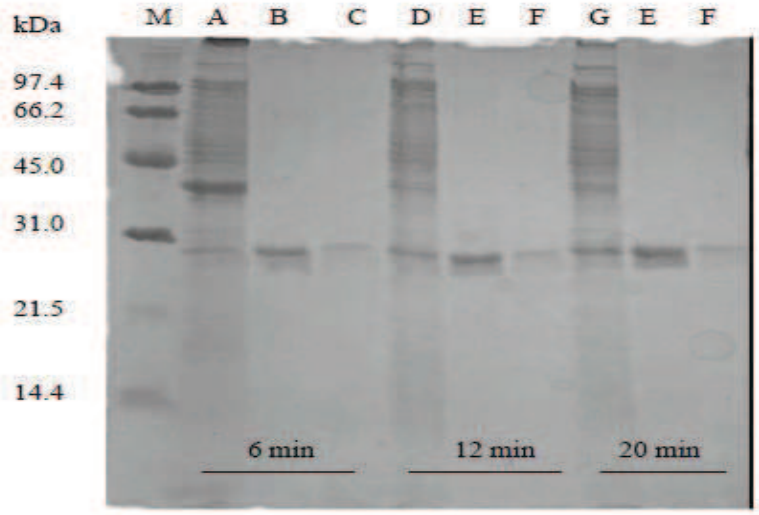

Figure 27: Gel image of Sonication optimization. M: molecular weight standard; A, D, G: supernatants; B, E, H: 75mM imidazole; C, F, I: 175mM imidazole.

Table 7: Relation between the sonication duration and the human p22HBP concentration after purification.

Sonication duration (min)	Concentration (µM)* ± 0.040
6	18.6
12	26.3
20	28.5

*V=3.5 mL

For the sonication testing, the SDS-PAGE image also shows protein bands at positions expected for human p22HBP. In the image we can see the increasing protein amount with the duration of the cell lysis. The gel results are confirmed with the protein concentration data after purification (**Table 7**). The protein concentration values suggest that there is a direct relation between the sonication duration and the amount of protein

extracted from the cells pellet, which is consistent with the fact that E. coli is not a good expression system in what concerns to its ability to release the heterologous proteins into the growth medium.

4.2. Human p22HBP structural studies

4.2.1. Overexpression and purification of Isotopic labeled human p22HBP

The labeling of human p22HBP samples was accomplished by growing the *E.* coli (pET-28a) cells in M9 medium with $^{15}NH_4Cl$, ^{13}C-glucose or both. Below (**Figure 27**) we can see the purification of a sample labeled with $^{15}NH_4Cl$ used for the studies of interaction with porphyrins. In the SDS-PAGE image we can see the protein expression profile after purification with an imidazole gradient in a NI-NTA column. In the gel we can see that the target protein band appears in a position expected for human p22HBP (more than 20 kDa). The gel also shows that the gradient of imidazole used (10-500 mM) was good because the protein was purified in a few steps and purest fractions were collected in the fraction 75 (mainly), 175 and 500 mM (do not appear in the gel) mM imidazole. These fractions were then gathered and concentrated. At right (**Figure 28**) the UV-Vis spectra of the protein is shown.

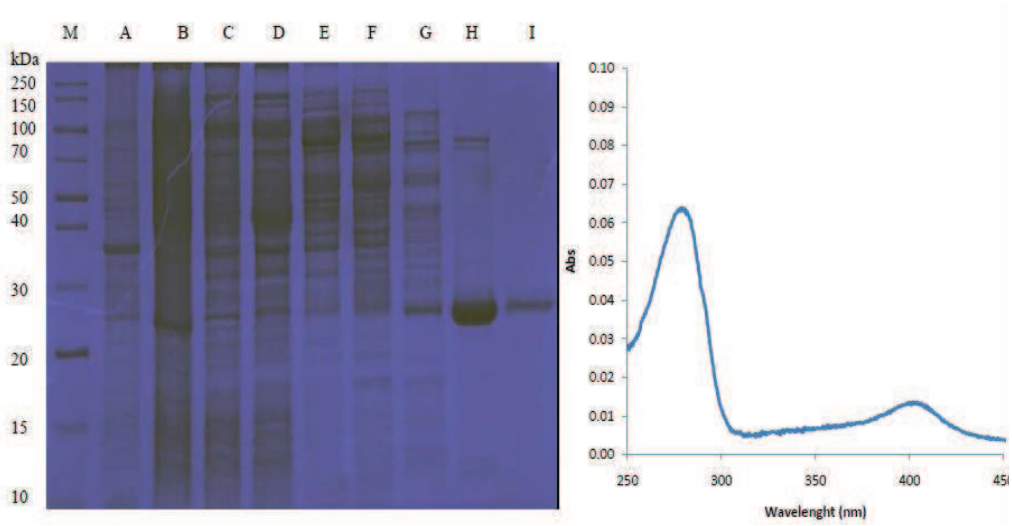

Figure 28: LEFT: Gel image of 15N human p22HBP. M: molecular weight standard; A to I: protein extract, pellet, 10, 20, 50, 75, 175, 500 mM. RIGHT: Absorbance of human p22HBP as a function of wavelength.

The concentration of human p22HBP is calculated by measuring the absorbance at 280 nm absorbance which is mainly a measurement of the number of Trp in the protein. The absorbance peak at 400 nm is caused by heme binded to the protein.

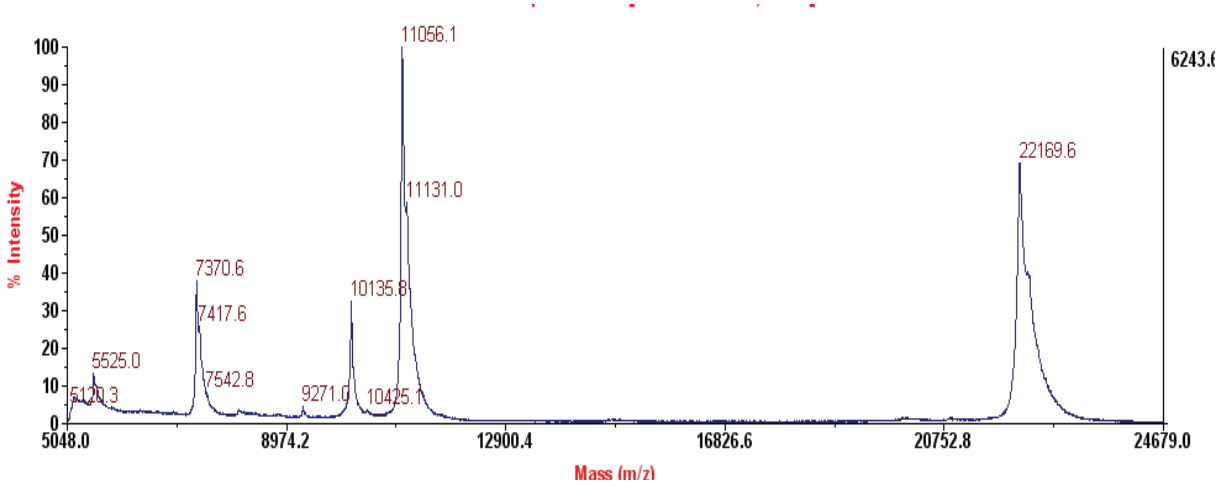

Figure 29: Mass spectra of human p22HBP.

The molecular mass of human p22HBP was accurately determined by mass spectra (**Figure 29**) which gave 22169.6 Da as expected for human p22HBP with a histidine tag. A theoretical molecular weight for human p22HBP was previously calculated using a biochemical protein parameter tool found in the Expasy (Bioinformatics Resource Portal) website based in the number and type of amino acids present in the sequence. The theoretical value was 22219 Da. The difference observed for the theoretical and experimental value is due to the fact that the protein was labeled.

4.2.2. 1D ^{1}H/2D ^{1}H,^{15}N NMR

The first NMR spectra acquired for human p22HBP was a 1D spectra which is very simple and fast to acquire. A good 1D NMR spectra usually means that the sample is good and we can use it to perform other kind of experiments (2D and 3D) which are the only way to extract important information (resonances of ^{1}H, ^{15}N and ^{13}C) for

biomolecular structure elucidation. A 1D spectrum of human p22HBP was acquired at 293K (**Figure 30**) using a 1 mM protein sample. We can see how crowded this spectrum is due to overlap of resonances.

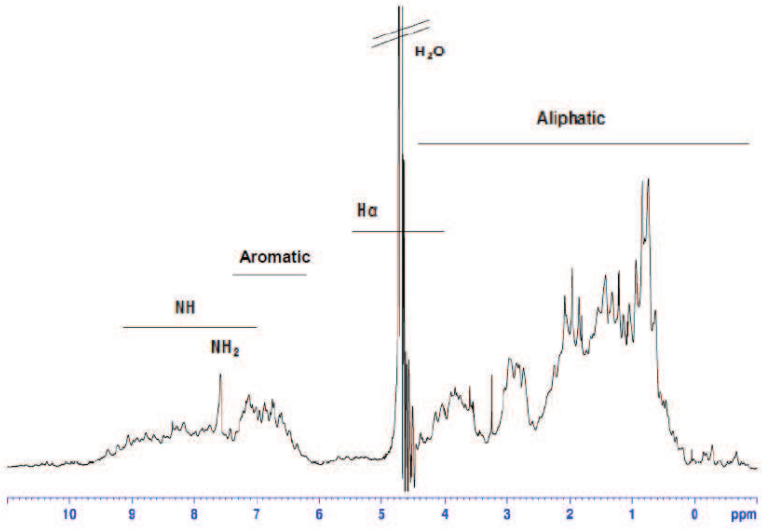

Figure 30: 1D spectrum of 1 mM human p22HBP acquired at 293K.

Being a relatively large protein in terms of NMR structure determination, using the conventional methods developed by Wuthrich we find two major problems: first, the large number of resonances causes signal overlap; second, the NMR signals relax faster, which leads to line broadening and poor sensitivity. To overcome this problem, the protein is labeled with $^{13}C/^{15}N$ and then pulse sequences based on the TROSY are used in the 2D/3D spectra to reduce transverse relaxation, thus obtaining sharper and more intense signals.

In **Figure 31**, an overlay of TROSY spectra of ^{15}N labeled human p22HBP acquired at 500 and 600 MHz is shown. This spectrum shows a lot of $^{1}H,^{15}N$ discrete cross-peaks detected, corresponding to the non-proline NH correlations from the amino acids in human p22HBP.

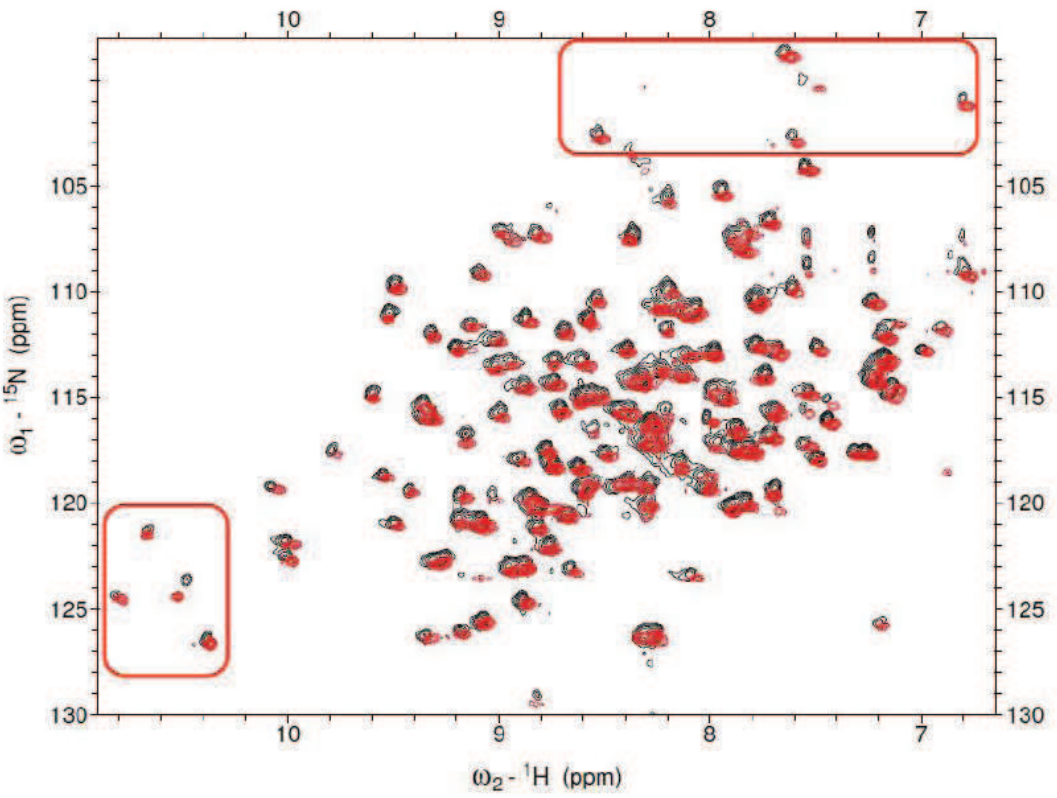

Figure 31: Superimposed ^{1}H,^{15}N TROSY spectrum of 1 mM human p22HBP recorded at 500 MHz (RED) and ^{1}H,^{15}N,^{13}C recorded at 600 MHz (**BLACK**) is shown. Both spectrum were acquired at 293 K in 50 mM phosphate/ 300 mM NaCl buffer at pH 8.0.

Both TROSY spectrum (**Figure 31**) of ^{15}N p22 HBP and ^{15}N/^{13}C human p22HB sample showed a good chemical shift dispersion which was indicative of a folded protein. The number of signals expected was 185 NH correlations but the analysis showed less than that number mainly because of the overlay. Analysis of both spectra showed that they can be superimposed. It is known that the higher the strength of the magnetic field applied, more sensitive NMR is in detecting the signals when using the same sample concentration. In the boxes: down - region of the Trp side-chain residues; up – backbone Gly. Both group of spins appears in a characteristic frequency range which allowed their identification by comparation with data from BMRB (**Table 2**). As necessary premise to calculate a structure by NMR, we carried out 3D heteronuclear experiment to assign the ^{1}H, ^{13}C and ^{15}N backbone chemical shifts of the protein. The assignment strategy will be described in the following section.

4.2.3. Sequence-specific resonance assignment

The sequence assignments of human p22HBP were obtained by a combination of approaches: triple resonance experiments, human/murine p22HBP structure alignment and porphyrin titration experiments. For the first strategy a labeled $^{15}N/^{13}C$ human p22HBP samples were used with a combination of 2D $^{1}H,^{15}N$ TROSY and 3D trHNCO, trHN(CO)CA and trHNCA (**Table 8**) spectra, which correlates the backbone chemical shifts to establish the sequential connectivity between amino acids.

Table 8: 3D experiments recorded and observed backbone chemical shift correlations.

TROSY-type experiments	Dimensionality	Observed correlation
HNCA	3D	$^{13}C^{\alpha}(i)$-$^{15}N(i)$-$^{1}H^{N}(i)$ $^{13}C^{\alpha}(i\text{-}1)$-$^{15}N(i)$-$^{1}H^{N}(i)$
HN(CO)CA	3D	$^{13}C^{\alpha}(i\text{-}1)$-$^{15}N(i)$-$^{1}H^{N}(i)$
HNCO	3D	$^{13}CO(i\text{-}1)$-$^{15}N(i)$-$^{1}H^{N}(i)$

The sequential backbone assignment was carried out semi-automatically using the program CARA. All peaks were picked manually. Generic spin systems, that are generated automatically when a peak is peaked, have been matched mainly using the TROSY-HNCA spectrum. If the signals of interest were present in the TROSY-HN(CO)CA, this spectrum was used for the distinction between C_i and C_{i-1}, otherwise this was determined from the relative signal intensities in the HNCA. Sequential peaks are normally less intense compared to intraresidue signals. Although the TROSY-HNCO is highly sensitive, it provides only one C'_{i-1} peak per residue and therefore does not allow the establishment of connectivities between generic spin systems. Generic spin systems which could be matched were manually connected leading to chains of various lengths. The obtained peptide segments were mapped to the protein sequence using a probability tool matching, which is an automatic strip matcher that gives the user all the best candidate for successors and predecessors of a specific system. The bigger the fragment is, less ambiguity the tool retrieves, being very helpful for fragments matching. Only unambiguous assignments were taken into account and a short sequential section is presented below (**Figure 32**).

In figure 29 we have slices from trHN(CO)CA and trHNCA spectra where we could assign the Cα resonances of human p22HBP (as shown in the following figure from Thr 43 to Val 46). The assignment using the triple resonance experiments is complemented with the comparation of the experimental chemical shift values with the average values from the BMRB database (**Table 2**).

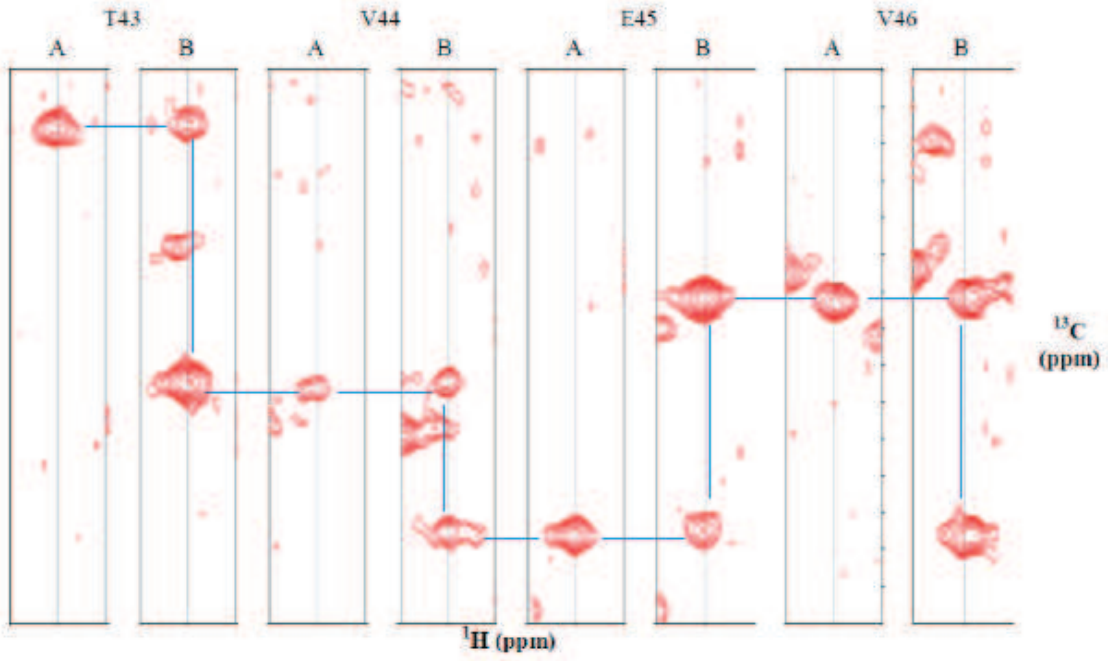

Figure 32: Sequential backbone assignments of the Thr 43 to Val 46 resonances for human p22HBP. A) trHN(CO)CA and B) trHNCA spectra acquired in the ^{15}N/^{13}C labeled human p22HBP.

The sequential walk was of great importance up to a certain point where 42 % of the residues in the human p22HBP sequence could be assigned. From that point on the task became difficult due to the increasing number of overlay peaks. To overcome this problem, we started using a second strategy based on the alignment of the two versions of p22HBP sequences: the human and the murine. If we perform a blast search in human and murine p22HBP using the blast tool from NCBI, we find that the first is a polypeptide with 189 amino acids and accession number NP_057071, and the second is a polypeptide with 190 amino acids and accession number NP_038574 (**Figure 33**).

Figure 33: Protein sequence alignment between murine and human p22HBP using CLUSTALW [40].

Looking at the alignment we can extract information such as the presence of 1 gap and the presence of 164/189 (87%) identical amino acids, which is a proof of the great level of homology between both polypeptides. Sharing this degree of homology, the resonance signals of conserved residues in both human and murine p22HBP should appear in the same region of a HSQC/TROSY spectrum.

The **Figure 34** shows the superposition of $^{1}H,^{15}N$ TROSY NMR spectrum of both proteins, acquired at 20ºC. In this spectrum we confirm the great level of homology in the two sequences as can be seen by the proximity of many murine resonances to human resonances, which was already expected. This strategy allows the identification of residues present in conserved sequences, without the need for triple resonance experiments (**Figure 35**). It can also be complementary to triple resonance experiments because it can be confirmed if a conserved residue was correctly assigned or not.

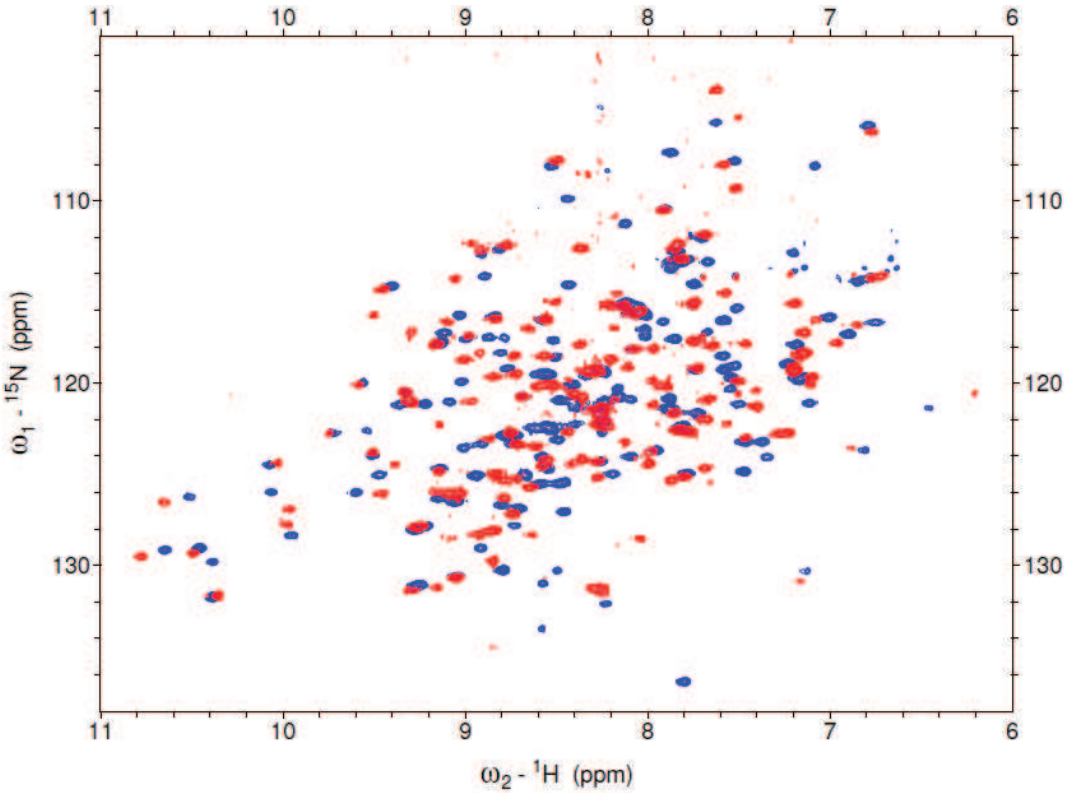

Figure 34: Superimposed TROSY spectra of murine p22HBP (BLUE) and human p22HBP (RED) at 20ºC.

The two strategies used until now, triple resonance experiments and murine/human p22HBP alignment, followed by overlay of the spectra, allowed the correct assignment of many signals but we still had many resonances that could not be assigned possibly due to proximity to non conserved residues or because of little or no conservation.

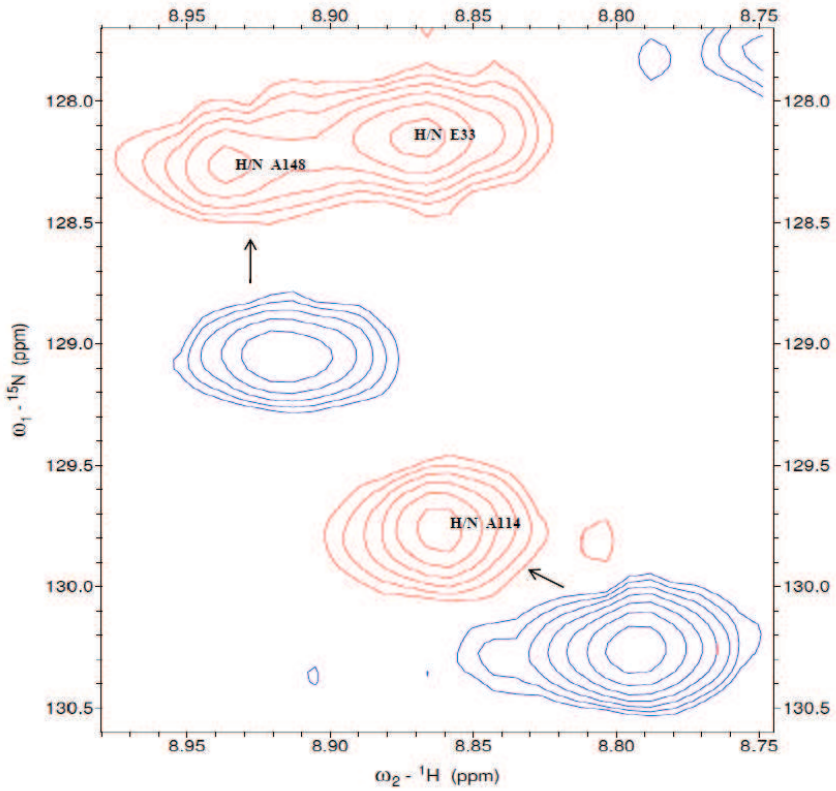

Figure 35: TROSY spectrum of human p22HBP (RED) and murine p22HBP (BLUE) showing the assignment of Ala 114 using the second strategy and Ala 148 using the third strategy.

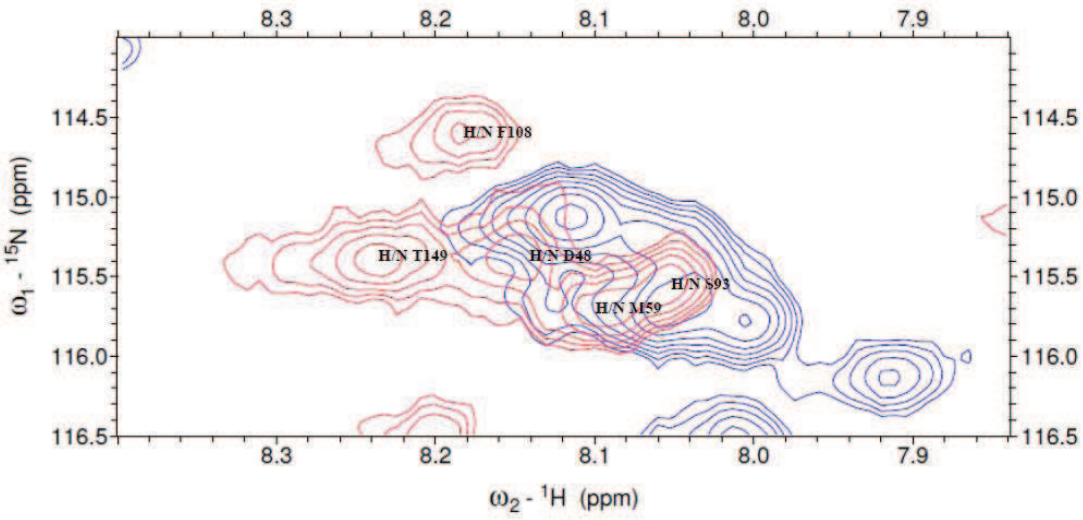

Figure 36: A region of the spectrum where the assignment using the second strategy was more difficult due to the overlap of the signals. RED: human p22HBP; BLUE: murine p22HBP.

Figure 35 shows how the overlay of murine with human p22HBP was helpful in assigning Ala 114. Looking at the alignment (**Figure 33**) Ala 114 is between prolines which has no signal in HSQC/TROSY spectrum so then the assignment of this residue could not be carried out with triple resonance experiments since we cannot link it to any other residue. However this residue is in a conserved sequence so then, superimposing the spectra of both proteins, we could see that the signals from Ala 114 for both proteins appear close to each other. In **Figure 36** we can see a region of the spectrum where the overlap of resonances made difficult to assign the spectrum. The assignment of the five residues was performed using triple resonance experiments (Asp 48 and Ser 93) and human/murine spectrum overlay (Met 59, Phe 108 and Thr 149). To assign further the backbone chemical shifts we used data from porphyrins titration studies that we will analyze in the following section.

4.3. Interaction studies

4.3.1. Chemical shift perturbation mapping

To better understand the binding mechanism to heme, NMR-monitored titrations were carried out to observe interactions between human p22HBP and hemin or PPIX. We acquired three spectra, human p22HBP:Tween, human p22HBP:Hemin and human p22HBP:PPIX. The sample with Tween was used as blank and the assignment was performed in order to eliminate the effect of chemical shift differences due to the presence of Tween. The binding kinetics that results in a slow exchange of protein conformational state, from free to bound, induces perturbation in sign and magnitude of the protein chemical shifts. This data is of great importance because the perturbations are higher in the nucleus near the binding pocket, so then finding these residues allows finding the place in the molecule where the interaction takes place. The analysis of the perturbation pattern presented by the residues, from free to bound state, allowed us also to make more assignments. As shown below for a region of spectra, the peaks from free to bound state presents a linear distribution (**Figure 37**).

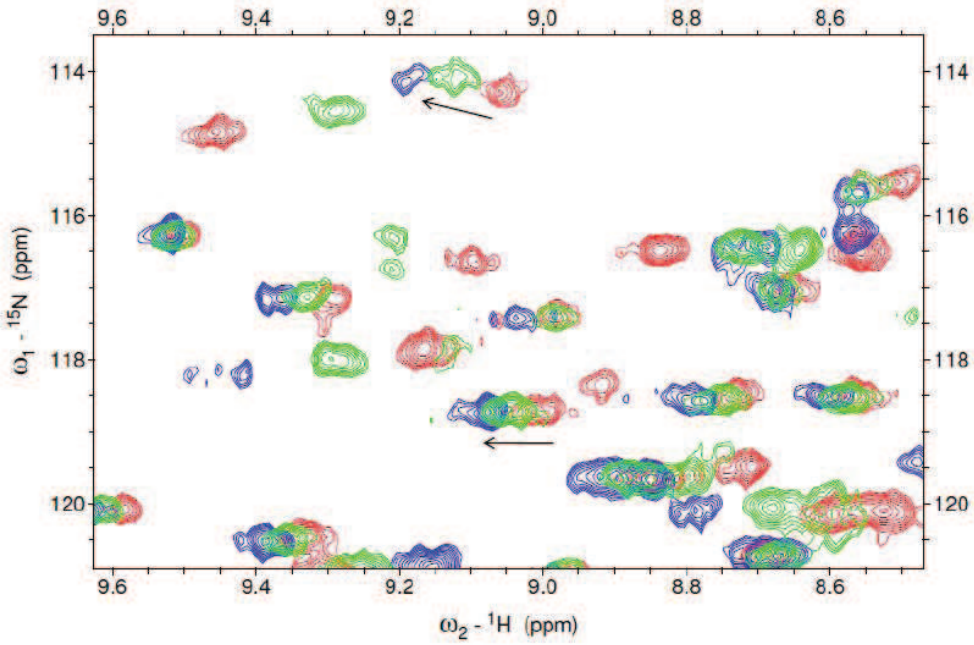

Figure 37: ^{1}H,^{15}N TROSY of human p22HBP titrated with porphyrins at 20ºC. RED-human p22HBP free, GREEN-human p22HBP titrated with PPIX, BLUE-human p22HBP titrated with hemin.

Figure 35 shows the overlay of the ^{15}N-labeled human p22HBP ^{1}H,^{15}N TROSY spectra recorded without (red) or with hemin (blue) and PPIX (green) (1:1.1 ratio). The first thing to notice is the dependency of protein chemical shifts on the nature of the ligand binding, which was expected because while the hemin is an iron containing porphyrins, PPIX is not. It has long been known that the effect of unpaired electrons on NMR spectral parameters such as chemical shifts and relaxation times is a valuable source of information on molecular structure. Electron-nuclear interactions involving a paramagnetic center with rapidly relaxing electron spins in an anisotropic ligand field give rise to large shifts in NMR frequencies [41]. The arrows (**Figure 37**) show the linear relation in chemical shift dispersion between the two porphyrins during binding with human p22HBP. The signal of NH group in HBP bound to PPIX appears closer to the signal of the NH group of the protein free and then the corresponding NH in hemin bound HBP.

The magnitude of the differences in chemical shift of the protein upon binding to the porphyrins was also studied. Significant chemical-shift changes were observed in several areas of the spectra. **Figure 38** shows the magnitude of the chemical-shift

change, for each assigned residue of the polypeptide. The biggest changes were seen for residues 42 to 56 and 139 to 179 which suggest that these residues participate in binding to porphyrins. In previous studies in murine p22HBP the residues found to have larger chemical shifts were 63, 139 and those in the region from 171-179 as can be seen in **Figure 38** [9].

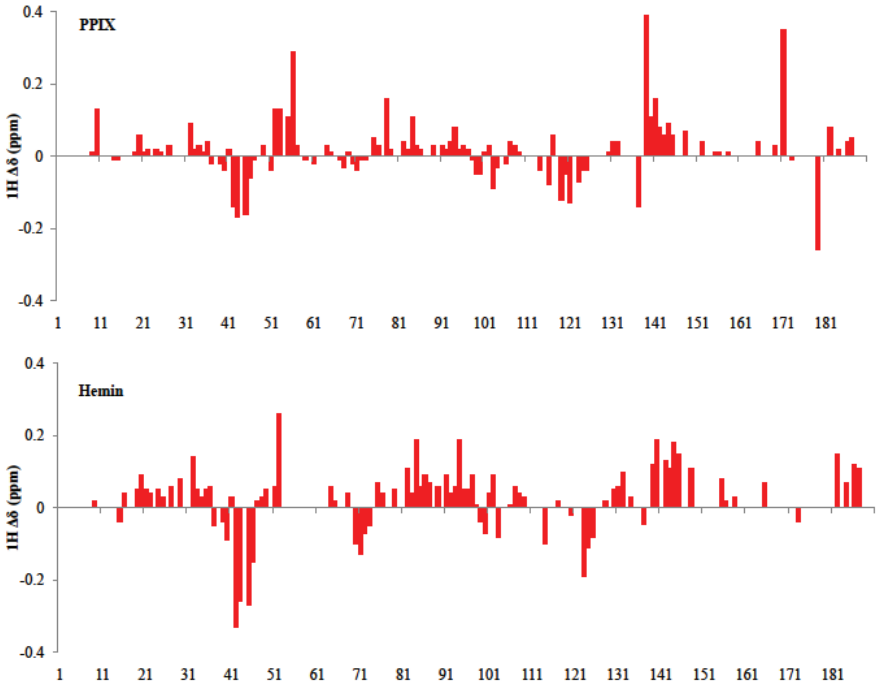

Figure 38: Experimental ^{1}H chemical shift differences observed for human p22 HBP upon binding of PPIX or Hemin plotted as a function of the human p22HBP amino acid sequence.

Figure 39: Comparation of the experimental ^{1}H chemical shift differences observed for human p22 HBP upon binding of PPIX or Hemin with the reported ^{1}H chemical shift differences observed for murine p22 HBP upon binding the same porphyrins.

Analyzing the overlay of both data (**Figure 39**) we can see that some of the residues with larger chemical shift change are the same (42 to 56 and 139 to 179), however, for murine p22HBP, 63 is one of the residues with larger chemical shift change, which is not concordant with the data obtained here, where the 63 almost does not shift at all. The plot shows that 63 appears in a region where few assignments were performed due to the overlap. In any case, the superposition of the murine/human titration data shows the same pattern in chemical shift differences which can mean that human and murine p22HBP interacts in the same way with these ligand and the same residues participate in this process. In the plot there are some regions far from the binding site that still do not present a chemical shift difference and these regions are the ones that we could not assign.

As referred previously, the analysis of the chemical shift differences data was of great utility mainly to confirm assignments but it also helped to assign Ala 148 (**Figure 35**) which could be assigned only using the data from the titration. The combination of all

the strategies mentioned back resulted in the assignment of 68% of the resonances (**Figure 40**, **41** and **Table A2** (Appendix)).

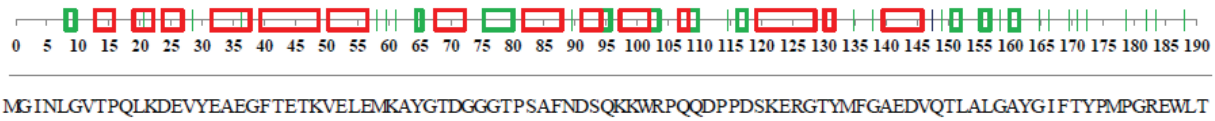

Figure 40: Human p22HBP residues assigned using the combination of: RED-triple resonance experiments (42%)); Green-sequence alignment (25%) corresponding to 68% and BLUE- titration data (0.5%).

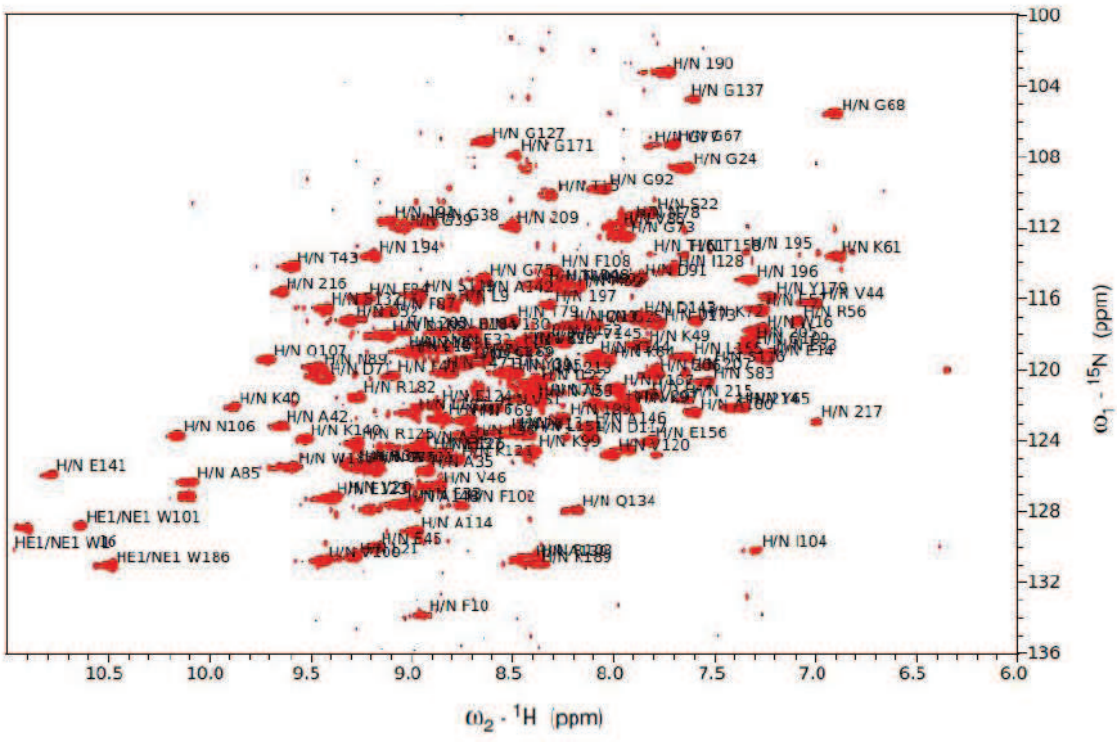

Figure 41: TROSY spectrum showing the assignment of 68 % of the human p22HBP residues.

4.3.2. Human p22HBP tetrapyrrole dissociation constants

In this section intrinsic fluorescence quenching was used to quantify binding constants for human p22HBP with hemin and PPIX. Recent studies performed in murine p22HBP reported values for the dissociation constants obtained for the PPIX and hemin complexes were 0.5 and 3.0 nM, respectively [9]. By using the same methodology we expect to get similar values for human p22HBP. The graph below (**Figure 42**) shows the intrinsic fluorescence quenching at maximum emission (340 nm) for human p22HBP as a function of increasing concentration of hemin and PPIX. Different concentrations (5 and 10 uM) of hemin and PPIX were tested and the dissociation constants were determined by non-linear fitting which gave the binding values that are represented in the following graph.

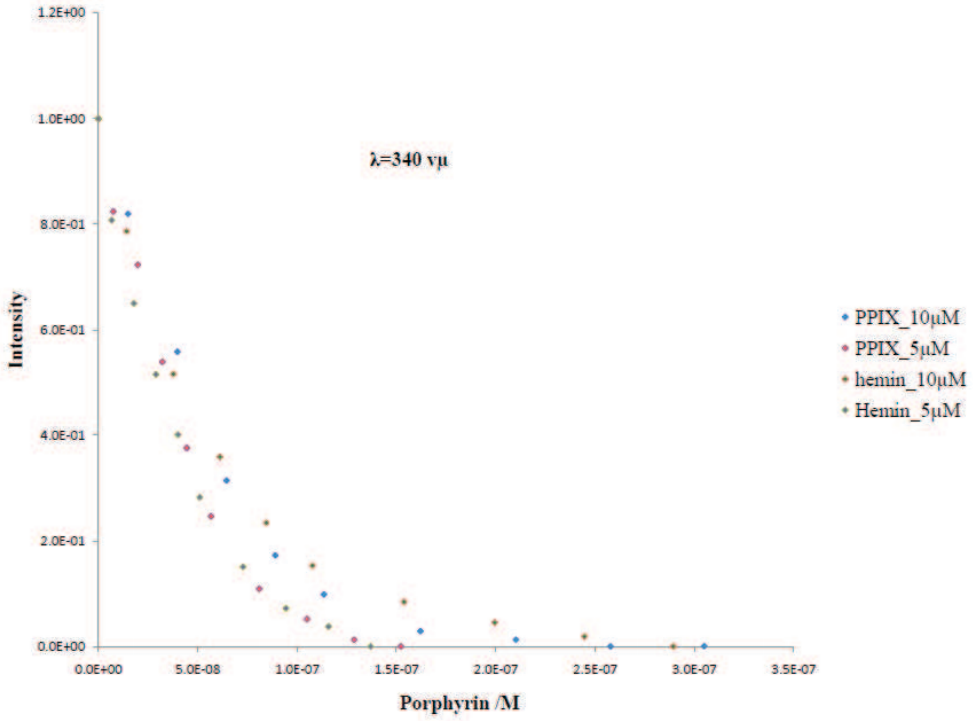

Figure 42: Intrinsic human p22HBP tryptophan fluorescence at maximum emission (340 nm) as a function of increasing PPIX or hemin concentrations.

In **Figure 42** it can be observed that each ligand caused efficient quenching of the p22 HBP tryptophan emission.

Porphyrin	Concentration /μM	Kd	ΔKd
PPIX	10	6.35E-09	1.32E-09
	5	4.7E-09	3.43E-09
Hemin	10	2.04E-08	4.60E-09
	5	2.55E-08	1.25E-08

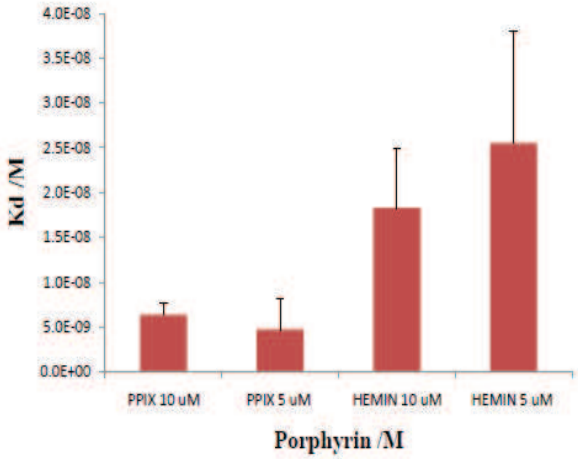

Figure 43: Dissociation constant values obtained for hemin and PPIX complexes.

Analyzing the data (**Figure 43**) we find that the values of dissociation constant obtained for the hemin and PPIX complexes were in the nanomolar range as reported previously for murine p22HBP. For the literature we know that even though the dissociation constant is higher for the hemin complex, the difference is not as pronounced as that obtained for human p22HBP here. Comparing with the literature, the PPIX presents a very similar dissociation constant, even though the concentration of the porphyrins is different (1.75 μM for hemin and 2.0 μM for PPIX) and it seems that the concentration of the porphyrins used is not so relevant for determining binding strength.

4.3.3. Separating human p22HBP from tetrapyrrole

In this section the separation of the human p22HBP from PPIX was tried with urea, which is widely used to assess protein unfolding [42]. After using a human:PPIX sample (1:1.1 mol ratio) for interaction studies, we wanted to separate the protein from the tetrapyrrole. First PPIX was added to the sample (~200 µL of human p22HBP:PPIX) to achieve 2.50 mL. Dialysis was carried out at 4ºC in this sample for 24 hours. After then the sample was diluted to 22.5 mL and loaded into a Ni-NTA column and the flow through was collected. Then, the column was washed first with 50 mM phosphate/300 mM imidazole buffer and then with the same buffer containing 175 mM imidazole. The eluates were recovered, concentrated and the protein and the PPIX concentration were quantified by UV-Vis and the results are listed below (**Table 9**). **Figure 44** shows the absorbance profile for the samples as a function of the wavelength. The protein absorbance peak appears at 280 nm while the PPIX appears at 400nm.

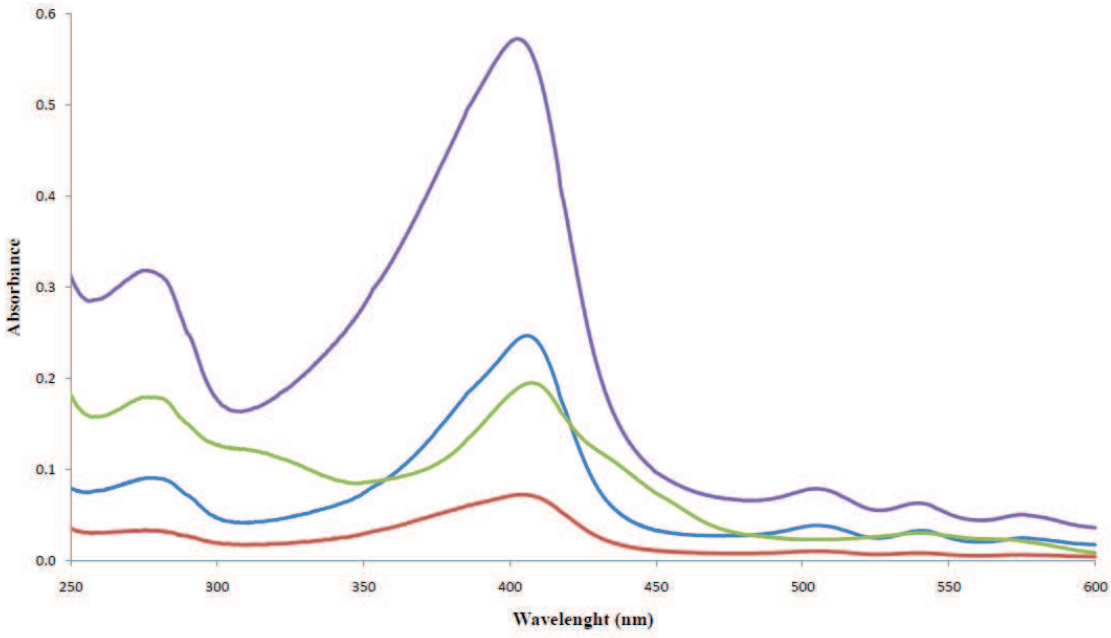

Figure 44: Absorbance profile of human p22HBP:PPIX complex after dyalisis and purification. Purple – initial sample, Red – 50 mM phosphate/300 mM NaCl buffer, Green – 50 mM phosphate/300 mM NaCl /175 mM imidazole, Blue – Flow through.

Table 9: Protein and PPIX concentrations for the different samples.

Samples	Concentration (μM)	
	Human p22HBP	PPIX
Initial	8.63	5.87
Flow through	2.46	2.43
Phosphate buffer	0.90	0.74
Phosphate buffer +175 mM imidazole	4.90	1.89

Analysis of the data (**Figure 44** and **Table 9**) shows that the purification column was not functioning correctly because the flow through and the phosphate buffer presented protein. When using this Ni-NTA column, for the protein concentration loaded, all the protein should complex with Ni^{2+} in the resin. Only the addition of a gradient of imidazole should break this interaction, which was not the case because when we washed the column with phosphate buffer the column also lost some protein. For PPIX, the data shows that the separation of the complex was not successful. The tetrapyrrole was washed always at same time with the protein. Although more experiments are needed and detailed experimental setups has to be carried out in order to determine if urea can separate the complex and if it is true, this may allow the reutilization of the protein samples.

Chapter 5 – General Discussion and Conclusions

In this work, in order to perform the proposed studies, the gene encoding the human p22HBP was cloned in the pET-28a plasmid, which was used to transform *E.* coli BL21(DE3) strains. To carry out structural studies by NMR, a good amount of quality sample is necessary, so then studies to optimize the parameters with influence in the protein yield are very important. In this work, the optimization of parameters (the time of induction, IPTG concentration and sonication duration) with influence in protein production were studied and the best results were obtained when inducing the target gene expression after 2 hours of growing the cells in M9 medium, using 0.1 mM of IPTG and 20 min of sonication to break the cells. For the first time, the sequence-specific resonance assignment of the human version of p22HBP was carried out using a combination of methods: triple resonance experiments (makes use of $^{15}N/^{13}C$ isotopic labeling), human/murine sequence and spectrum at 293K overlay and finally the data from interaction studies with tetarpyrroles. This resulted in assignment of 68% of the NH cross peaks present in the $^{1}H,^{15}N$ TROSY spectrum. All these methods were important because in certain regions of the spectrum, the resonances overlap made difficult the assignment and we could choose the best one to use in that moment. In future the assignment can be totally concluded labeling the protein with $^{15}N/^{13}C/^{2}H$ which combed to HNCACB will decrease the resonances overlap allowing the spins to be easily assigned. During preparation and storage of the proteins, we notice that the buffer can cause a fast protein precipitation when a protein sample containing 50 mM phosphate/ 300 mM NaCl is stored at -20ºC. To store this sample, a desalting in PD-10 column is required, in order to eliminate the NaCl.

Analysis of the chemical shift changes of human and murine p22 HBP in upon tetrapyrrole binding indicates that the same residues are present in the binding pocket, so then they are important in the way the protein interacts with the tetrapyrroles. In addition, the human p22HBP binding affinities were determined and they suggests that they are in the nanomolar range as reported for murine p22HBP, which suggest that the porphyrins binds with high affinity even though the difference in the values obtained for PPIX and hemin. This difference may be related to errors during the experimental procedure, so then more fluorescence quenching studies can be performed in order to assess better conclusions. Finally, an attempt was carried out to separate the protein from the PPIX. For that a dialysis was performed and after then a purification in Ni-NTA column was carried out and the results showed that this temptative failed.

However the results suggests that the column used was not working well, which means that more studies have to be performed in order to assess this conclusion. The advantage of the complex separation is the reutilization of the protein sample avoiding the continuously over-expressions.

Chapter 6 – References

1 Wijayanti, N., Katz, N. & Immenschuh, S. Biology of heme in health and disease. *Curr. Med. Chem.* **11**, 981-986 (2004).
2 Cunha, L. F. L. Genetic, biochemical and biophysical studies of the heme biosynthetic pathway phD thesis, *Universidade do Porto* (2008).
3 Ajioka, R. S., Phillips, J. D. & Kushner, J. P. Biosynthesis of heme in mammals. *Biochim. Biophys. Acta-Mol. Cell Res.* **1763**, 723-736 (2006).
4 Taketani, S. *et al.* Molecular characterization of a newly identified heme-binding protein induced during differentiation of urine erythroleukemia cells. *Journal of Biological Chemistry* **273**, 31388-31394 (1998).
5 Zylka, M. J. & Reppert, S. M. Discovery of a putative heme-binding protein family (SOUL/HBP) by two-tissue suppression subtractive hybridization and database searches. *Molecular Brain Research* **74**, 175-181 (1999).
6 Sato, E. *et al.* SOUL in mouse eyes is a new hexameric heme-binding protein with characteristic optical absorption, resonance Raman spectral, and heme-binding properties. *Biochemistry* **43**, 14189-14198 (2004).
7 Blackmon, B. J., Dailey, T. A., Xiao, L. C. & Dailey, H. A. Characterization of a human and mouse tetrapyrrole-binding protein. *Archives of Biochemistry and Biophysics* **407**, 196-201 (2002).
8 Babusiak, M., Man, P., Sutak, R., Petrak, J. & Vyoral, D. Identification of heme binding protein complexes in murine erythroleukemic cells: Study by a novel two-dimensional native separation - liquid chromatography and electrophoresis. *Proteomics* **5**, 340-350 (2005).
9 Dias, J. S. *et al.* The first structure from the SOUL/HBP family of heme-binding proteins, murine P22HBP. *Journal of Biological Chemistry* **281**, 38966-38966 (2006).
10 Micaelo, N. M., Macedo, A. L., Goodfellow, B. J. & Félix, V. Tetrapyrrole binding affinity of the murine and human p22HBP heme-binding proteins. *Journal of Molecular Graphics and Modelling* **29**, 396-405 (2010).
11 Qoronfleh, M. W., Hesterberg, L. K. & Seefeldt, M. B. Confronting high-throughput protein refolding using high pressure and solution screens. *Protein Expression and Purification* **55**, 209-224, (2007).
12 Caruso, C. S. Clonagem, Expressão e Caracterização de Proteínas Recombinantes de Xylella Fastidiosa PhD thesis, *Universidade de São Paulo* (2007).
13 Itakura, K., T. Hirose, R. & Riggs, A. D. Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin. *Science* **198**, 1056-1063 (1977).
14 Schmidt, F. R. Recombinant expression systems in the pharmaceutical industry. *Applied Microbiology and Biotechnology* **65**, 363-372, (2004).
15 Rai, M. & Padh, H. Expression systems for production of heterologous proteins. *Current Science* **80**, 1121-1128 (2001).
16 Schumann, W. & Ferreira, L. C. S. Production of recombinant proteins in Escherichia coli. *Genetics and Molecular Biology* **27**, 442-453 (2004).
17 Kraljevic, J. B. & Mrzljak, G. P. Supramolecular Structure and Function 10. *Springer* (2011).
18 Makrides, S. C. Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiological Reviews* **60**, 512-& (1996).
19 Novagen. pET System Manual (2011).
20 Dingermann, T., Steinhilber, D., folkers, G. Methods and Principles in Medicinal Chemistry Vol. 21. *Wiley* (2004).

21 Arnau, J., Lauritzen, C., Petersen, G. E. & Pedersen, J. Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expression and Purification* **48**, 1-13 (2006).
22 QIAGEN. A handbook for high-level expression and purification of 6xHis-tagged proteins (2003).
23 http://www.enciclopedia.com.pt/print.php?type=A&item_id=1038 (2011).
24 Wuthrich, K. NMR of Proteins and Nucleic Acids. *Wiley* (1986).
25 Kendrew, J. C. *et al.* Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å. Resolution. *Nature* **185** (1960).
26 Koskela, O. NMR Spectroscopy of multi-domain proteins:immunoglobulin-like domains of human filamin A phD thesis, *University of Helsinki* (2009).
27 Kai, L. Development of computational methods for the rapid determination of NMR resonance assignment of large proteins phD thesis, *National University of Singapore* (2003).
28 Teng, Q. Structural Biology: Practical NMR Applications. *Springer* (2005).
29 Evans, J. N. S. Biomolecular NMR Spectroscopy. Vol. 70. *Oxford university press* (1996).
30 Pervushin, K. Impact of Transverse Relaxation Optimized Spectroscopy (TROSY) on NMR as a technique in structural biology. *Quarterly Reviews of Biophysics* **33**, 161-197 (2000).
31 Schubert, M. Development and application of novel triple-resonance experiments for the assignment of protein NMR spectra (1998).
32 Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* **34**, 93-158 (1999).
33 Goto, N. K. & Kay, L. E. New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Current Opinion in Structural Biology* **10**, 585-592 (2000).
34 Salzmann, M., Wider, G., Pervushin, K., Senn, H. & Wuthrich, K. TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins. *J. Am. Chem. Soc.* **121**, 844-848 (1999).
35 Fernandez, C. & Wider, G. TROSY in NMR studies of the structure and function of large biological macromolecules. *Current Opinion in Structural Biology* **13**, 570-580 (2003).
36 Meyer, B. & Peters, T. NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angewandte Chemie-International Edition* **42**, 864-890 (2003).
37 Clarkson, J. & Campbell, I. D. Studies of protein-ligand interactions by NMR. *Biochemical Society Transactions* **31**, 1006-1009 (2003).
38 Nienhaus, G. U. Methods in Molecular Biology. *Humana Press* (2005).
39 Lakowicz, J. R. Principles of Fluorescence Spectroscopy. *Plenum* (1999).
40 Thompson, J. D., Higgins, D. G. & Gibson, T. J. Clustal-w - Improving the sensitivity of progressive multiple sequence alignmnent through sequence weighting, position-specific gap penalties and weight matrix choice *Nucleic Acids Res.* **22**, 4673-4680 (1994).
41 Gochin, M. & Roder, H. Protein-structure refinement based on paramagnetic NMR shifts - Applications to wild- type and mutant forms of cytochrome-C. *Protein Science* **4**, 296-305 (1995).

42 Bennion, B. J. & Daggett, V. The molecular basis for the chemical denaturation of proteins by urea. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 5142-5147 (2003).

Chapter 7 – Appendix

Table A 1: Fluorescence quenching measurements for Kd determination of the human p22HBP(100 nm) – PPIX (10 μM) complex.

Intensity (340nm)	V_{HEMIN} ad. (L)	V_{HEMIN} ad. total(L)	Dil. factor	Int.(340nm) (corr)	$n_{adic.}$ /mol	$[PPIX]_{cell}$ /M	$[soul]_{cell}$ /M	Normalization
3.87E+06	0.00E+0	0.00E+0	1.00E+0	3.87E+06	0.00E+0	0.00E+0	1.00E-07	1.00E+0
3.38E+06	3.00E-06	3.00E-06	1.00E+00	3.39E+06	3.00E-11	1.50E-08	9.99E-08	8.20E-01
2.69E+06	5.00E-06	8.00E-06	1.00E+00	2.70E+06	7.99E-11	3.98E-08	9.96E-08	5.58E-01
2.04E+06	5.00E-06	1.30E-05	1.01E+00	2.05E+06	1.30E-10	6.45E-08	9.94E-08	3.14E-01
1.66E+06	5.00E-06	1.80E-05	1.01E+00	1.67E+06	1.80E-10	8.91E-08	9.91E-08	1.73E-01
1.46E+06	5.00E-06	2.30E-05	1.01E+00	1.48E+06	2.30E-10	1.14E-07	9.89E-08	9.86E-02
1.27E+06	1.00E-05	3.30E-05	1.02E+00	1.29E+06	3.30E-10	1.62E-07	9.84E-08	2.91E-02
1.22E+06	1.00E-05	4.30E-05	1.02E+00	1.25E+06	4.30E-10	2.10E-07	9.79E-08	1.27E-02
1.19E+06	1.00E-05	5.30E-05	1.03E+00	1.22E+06	5.29E-10	2.58E-07	9.74E-08	0.00E+00
1.18E+06	1.00E-05	6.30E-05	1.03E+00	1.22E+06	6.29E-10	3.05E-07	9.69E-08	3.40E-04

Table A 2: Values of chemical shits for the nucleus assigned in the human p22HBP.

Residue	$\delta\ ^{1}H^{N}$ [ppm]	$\delta\ ^{15}N$ [ppm]	$\delta\ ^{13}CO$ [ppm]	$\delta\ ^{13}C\alpha$ [ppm]
1				
2				
3				
4				
5				
6				
7				
8				
9	8.79	116.4	173.9	55
10	8.92	134.02	171.364	56.28
13	8.48	123.48		54.81
14	7.23	119.44	170.419	56.93
15				
16				
17				
18				
19	8.22	117.48	173.907	50.89
20	9.33	127.12	172.766	60.7
21	9.28	130.6	173.386	53
22	7.81	111.2	174.820	55.71
23				
24	7.63	108.66	172.147	41.91
25	7.32	118.34		63.27
26				59.18
27	7.83	121.37	172.701	55.04
28				
29	7.95	121.97	173.027	59.66
30				
31	8.56	120.13	175.570	52.43
32	8.85	118.89	170.745	51.67
33	8.98	127.51	171.346	51.41

34	9.18	125.48	172.897	50.86
35	8.91	125.66	171.038	47.95
36	9.21	125.41	174.625	54.76
37	8.88	124.68	172.831	50.86
38	8.91	111.83		41.48
39	9.03	112.08	169.375	41.15
40	9.87	122.14	169.310	53.04
41	9.09	120.33	172.962	53.76
42	9.64	123.16	171.918	47.48
43	9.59	114.23		54.52
44	6.99	116.25	170.940	58.52
45	9.17	130.08	171.332	52.24
46	8.86	126.61	172.342	58.71
47	8.85	120.08	178.668	58.23
48	8.25	115.23	170.549	53.85
49	7.85	118.56	178.276	50.71
50				
51	8.59	122.2		65.13
52	9.3	117.21	173.777	56.27
53	7.22	119.05	170.940	55.65
54	8.95	124.29	176.255	52.38
55	8.61	119.45	175.896	55.38
56	7.08	117.2	176.124	55.85
57	7.27	116.58	116.576	54.85
58	8.34	121.66	170.386	47.57
59	8.2	115.58	173.092	54.85
60				
61	6.9	113.61	175.961	57.14
62				
63				
64	8.02	119.49	174.396	54.04
65	7.35	122.11	174.331	55.66
66				
67	7.7	107.35	173.190	41.63
68	6.89	105.59		41.77
69	8.72	122.83	171.788	58.28
70	8.4	121.61	170.810	49.24
71	9.43	120.47	175.916	54.28
72	7.58	117.21	176.222	52.09
73	7.94	112.53	173.744	43.62
74				
75	8.64	114.93	174.233	52.33
76	8.84	122.75	172.473	53.28
77	7.81	107.46	175.798	45.1
78	7.91	111.68	168.527	50.95
79	8.69	117.87	175.048	54
80				
81				
82	8.47	118.75		51.76
83	7.53	120.67	174.527	51.33
84	9.23	116.06	170.419	53
85	10.08	126.32	169.832	48.24
86	7.98	112	172.831	56.28
87	9.1	116.78	170.940	52.56
88				
89	9.45	119.87	173.744	49.61
90				
91	7.87	114.94	176.450	50.51
92	8.04	109.85	174.331	42.37
93	8.16	115.42	171.038	55.84

94	8.72	119.61	173.069	58.88
95	8.49	120.22	175.407	53.28
96	8.42	118.73	176.222	54.74
97	7.91	122.02	175.505	53.71
98	8.71	123.89	170.549	52.6
99	8.39	124.55	172.864	53.32
100	9.43	130.8	172.473	60.74
101	9.28	125.33	172.833	53.22
102	8.75	127.64	171.918	51.94
103	8.38	130.65	170.614	47.75
104	7.29	130.16	171.788	55.65
105				
106	10.15	123.72	173.875	54.32
107	9.7	119.41	175.570	55.31
108	8.28	114.46	171.560	57.72
109	7.32	118.71	171.560	56.55
110				
111				
112				
113				
114	8.97	129.13	173.125	47.46
115				
116	7.52	119.79	176.907	56.84
117	8.11	123.78	176.190	52.7
118				
119	8.96	115.85	174.429	57.55
120	8	124.74	170.093	58.22
121	8.77	125.07	171.429	50.94
122	8.38	120.81	170.777	58.26
123	9.39	127.24	174.233	52.32
124	8.87	121.98	171.429	52.13
125	9.27	124.13	173.157	53.32
126	8.91	124.69	171.592	53.22
127	8.62	107.19	173.712	43.66
128	7.7	114.43	170.516	56.69
129	8.65	121.11		61.21
130	8.49	117.23	170.712	58.03
131	8.97	119	171.718	54.13
132	9.42	116.51	171.788	53.08
133				
134	8.17	127.91	172.440	55.31
135				
136				
137	7.64	104.76	171.984	41.39
138				
139	8.41	130.73	172.701	49.46
140	9.52	123.89	173.288	51.27
141	10.78	125.88	174.559	59.22
142	8.68	115.9	174.266	52.32
143	7.87	117.01	178.309	54.36
144	8.09	119.22	185.417	59.09
145	8.24	118.42	178.021	63.65
146	8.12	123.19	176.255	52.75
147				
148	9.05	127.62	171.103	51.65
149	8.34	115.25	172.440	41.61
150				
151	8.42	123.68	170.810	50.08
152	8.33	118.22	172.407	56.61
153				

154				
155	7.63	119.29	181.048	52.27
156	7.81	124.02	170.419	59.93
157				
158	7.64	113.56	177.298	62.83
159				
160	7.59	122.43	176.222	58.07
161				
162				
163				
164				
165	9.11	118.04	170.745	51.51
166				
167	7.97	121.08	172.994	51.03
168				
169	8.67	119.46	171.429	52.6
170				
171	8.48	107.92	172.994	43.09
172				
173	7.77	117.41	172.440	56.69
174				
175				
176				
177				
178				
179	7.21	115.94	173.679	52.84
180				
181				
182	9.26	121.51	172.310	53.89
183				
184	8.85	117.86	178.635	55.03
185				
186	9.58	125.51	171.690	50.27
187	8.99	122.44	173.255	50.41
188				
189	8.38	131.05	170.647	49.46